

# The Modified Stochastic Games\*

Eilon Solan<sup>†</sup>

March 14, 2017

## Abstract

We present a new tool for the study of multiplayer stochastic games, namely the modified game, which is a normal-form game that depends on the discount factor, the initial state, and for every player a partition of the set of states and a vector that assigns a real number to each element of the partition. We study properties of the modified game, like its equilibria, min-max value, and max-min value. We then show how this tool can be used to prove the existence of a uniform equilibrium in a certain class of multiplayer stochastic games.

**Keywords:** Stochastic games, modified game, uniform equilibrium.

**JEL Classification:** C72, C73.

## 1 Introduction

Stochastic games is a model for studying dynamic models in which the state variable changes in response to the players' actions. Shapley (1953) presented the

---

\*The author thanks Eitan Altman for helping in identifying relevant references, Omri Solan for helpful discussions, and acknowledges the support of the Israel Science Foundation, Grant #323/13.

<sup>†</sup>The School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997800, Israel. e-mail: eilons@post.tau.ac.il

model of two-player zero-sum discounted stochastic games with finitely many states and actions and proved the existence of the value and of stationary optimal strategies for the two players. The main approach for studying discounted stochastic games is using their recursive structure. This approach was successfully utilized by Fink (1964) and Takahashi (1964) to prove the existence of a discounted equilibrium in stationary strategy in multiplayer stochastic games, by Hörner, Sugaya, Takahashi, and Vieille (2011) for characterizing the limit set of discounted equilibrium payoffs in a certain class of multiplayer stochastic games, and by, e.g., Mertens and Parthasarathy (1987) to study equilibrium in multiplayer stochastic games with general state and action spaces.

Mertens and Neyman (1981) suggested a robust equilibrium concept for stochastic games, namely uniform  $\varepsilon$ -equilibrium. A strategy profile is a uniform  $\varepsilon$ -equilibrium if it is an  $\varepsilon$ -equilibrium in the discounted game, provided the players are sufficiently patient, and in the finite-horizon game, provided the game is sufficiently long. The study of uniform  $\varepsilon$ -equilibrium and its variants, namely various types of uniform correlated  $\varepsilon$ -equilibrium, turned out to be difficult, and various techniques were used in the literature by different authors. These include the vanishing discount factor approach (Vrieze and Thuijsman, 1989), graph theoretic tools (Vieille, 2000a,b), the introduction of a modified game (Solan, 1999, 2000; Solan and Vohra, 2002), a dynamical system approach (Solan and Vieille, 2001), and a topological approach (Simon, 2007, 2012).

In the present paper we extend to general stochastic games the approach taken by Solan (1999, 2000) and Solan and Vohra (2002). These papers study multiplayer absorbing games,<sup>1</sup> and define a *modified* game, in which the stage payoff in the nonabsorbing state is different than the stage payoff in the original game: the new stage payoff of each player is the minimum between his *expected* stage payoff (given the mixed actions chosen by the players) and the uniform min-max value of the player in the nonabsorbing state. It is shown that this game admits a discounted equilibrium in stationary strategies and that the limit of the discounted equilibrium

---

<sup>1</sup>A state in a stochastic games is *absorbing* if the play cannot leave it once this state was reached. A stochastic game is *absorbing* if all states but one are absorbing.

payoffs is at least the uniform min-max value of the original game. Finally, like in the vanishing discount factor approach, the sequence of discounted stationary equilibria is used to construct a uniform  $\varepsilon$ -equilibrium in three-player absorbing games and in team games, and a uniform normal-form correlated  $\varepsilon$ -equilibrium in multiplayer absorbing games.

The fact that in absorbing games there is a single nonabsorbing state simplified the definition and the study of the modified game, as the expected stage payoff and the uniform min-max value of the players are those given in the nonabsorbing state. In general stochastic games, where the state changes from stage to stage, it is not clear how to define the payoff function in the modified game so as to keep the useful properties that the definition of the modified game for absorbing games has.

Below we provide a definition for the modified game that retains the features that were valuable in earlier studies. The modified game is a normal-form game that depends on the initial state, a discount factor, and for each player  $i$  a partition  $\mathcal{D}^i$  of the set of states and a collection of cut-offs ( $c^i(D)$ ), one for every element  $D$  of the partition  $\mathcal{D}^i$ . In this game, each player chooses a strategy in the stochastic game, and his payoff is the sum over all elements  $D$  of the partition, of the minimum between his expected discounted payoff restricted to stages spent in  $D$  and his cut-off  $c^i(D)$ . Thus, the expected discounted payoff of a player  $i$  in an element  $D$  of his partition cannot be higher than the cut-off  $c^i(D)$ .

We then study properties of the modified game. We show that it admits an equilibrium in stationary strategies, and we compare its min-max value and max-min value, both in general strategies and in stationary strategies, to the min-max value and max-min value of the original game. In particular we show that if the partitions ( $\mathcal{D}^i$ ) satisfy a certain property and the cut-offs ( $c^i(D)$ ) are not too low, then the limit of a sequence of equilibrium payoffs in the modified game as the discount factor goes to 1 (patience) is at least the uniform min-max value of the initial state in the original stochastic game. We finally provide an application of this tool to the study of uniform  $\varepsilon$ -equilibrium in a certain class of multiplayer stochastic games.

In addition to providing a new tool to study multiplayer stochastic games, the paper shows that to study stochastic games it may be useful to define auxiliary games and study their properties. The modified game that we present is just one possible auxiliary game.

The paper is organized as follows. In Section 2 we describe the model of stochastic games. In Section 3 we define the modified game and summarize the results that are proven in the rest of the paper. In Section 4 we study equilibria of the modified game. In Section 5 we present the notion of uniform max-min value in stochastic games. The max-min value and the min-max value of the modified game, and their comparison to the uniform max-min value and uniform min-max value in the original game, are studied in Sections 6 and 7 respectively. The min-max and max-min values in stationary strategies of the modified game are studied in Sections 8 and 9 respectively. The application to the uniform equilibrium appears in Section 10.

## 2 Stochastic Games: the Model

### 2.1 The Game Play

A multiplayer *stochastic game* is a vector  $\Gamma = (I, S, (A^i)_{i \in I}, (u^i)_{i \in I}, q)$  where

- $I = \{1, 2, \dots, |I|\}$  is a finite set of players.
- $S$  is a finite set of states.
- $A^i(s)$  is a finite set of actions available to player  $i$  in state  $s$ . Denote by  $A(s) := \times_{i \in I} A^i(s)$  the set of all action profiles available at state  $s$ . Denote by  $\Lambda := \{(s, a) : s \in S, a \in A(s)\}$  the set consisting of pairs of state and action profile available at that state.
- $u^i : \Lambda \rightarrow \mathbb{R}$  is player  $i$ 's payoff function. We assume w.l.o.g. that the payoffs are bounded between -1 and 1.

- $q : \Lambda \rightarrow \Delta(S)$  is a transition function, where  $\Delta(X)$  is the set of probability distributions over  $X$ , for every nonempty finite set  $X$ .

The game is played as follows. The initial state  $s_0 \in S$  is given. At each stage  $n \in \mathbb{N} \cup \{0\}$ , the current state  $s_n$  is announced to the players. Each player  $i \in I$  chooses an action  $a_n^i \in A^i(s_n)$ ; the action profile  $a_n = (a_n^i)_{i \in I}$  is publicly announced,  $s^{n+1} \in S$  is drawn according to  $q(\cdot \mid s_n, a_n)$ , and the game proceeds to stage  $n + 1$ .

We extend the domain of  $q$  and  $(u^i)_{i \in I}$  to  $\cup_{s \in S} (\{s\} \times \Delta(A(s)))$  in a multilinear fashion: for every state  $s \in S$  and every mixed action profile  $\alpha \in \times_{i \in I} \Delta(A^i(s))$  we define

$$q(s, \alpha) := \sum_{a \in A(s)} \alpha[a] q(s, a),$$

and

$$u^i(s, \alpha) := \sum_{a \in A(s)} \alpha[a] u^i(s, a),$$

where  $\alpha[a] := \prod_{i \in I} \alpha^i(a^i)$ .

## 2.2 Strategies and Payoffs

A *finite history of length  $n$*  is a sequence  $h_n = (s_0, a_0, \dots, s_n) \in \Lambda^n \times S$ . By convention, the set  $\Lambda^0$  contains only the empty history. Let  $H := \cup_{n \geq 0} (\Lambda^n \times S)$  be the set of all finite histories and  $H^\infty := \Lambda^\infty$  be the set of *plays*. When  $h \in H^\infty$  is a play and  $n \geq 0$  we denote by  $h_n \in \Lambda^n \times S$  the prefix of  $h$  of length  $n$ . The space  $H^\infty$  together with the  $\sigma$ -algebra generated by all finite cylinders is a measure space. We denote by  $\mathcal{H}(n)$  the algebra generated by the finite histories of length  $n$ . We assume perfect recall. Accordingly, a (behavior) *strategy* of player  $i$  is a function  $\sigma^i : H \rightarrow \Delta(A^i)$  such that  $\sigma^i(h_n) \in \Delta(A^i(s_n))$  for every finite history  $h_n = (s_0, a_0, \dots, s_n) \in H$ . Denote by  $\Sigma^i$  the set of all strategies of player  $i$ , by  $\Sigma := \times_{i \in I} \Sigma^i$  the set of all strategy profiles, and by  $\Sigma^{-i} := \times_{j \neq i} \Sigma^j$  the set of all strategy profiles of all players except player  $i$ .

A class of simple strategies is the class of *stationary strategies*. Those are strategies in which the choice of a player at each stage depends only on the current state, and not on previously visited states or on past choices of the players. A stationary strategy of player  $i$  can be identified with an element of  $\Sigma_{\text{stat}}^i := \times_{s \in S} \Delta(A^i(s)) \subset \mathbb{R}^{\sum_{s \in S} |A^i(s)|}$ , and will be denoted  $x^i = (x^i(s))_{s \in S}$ . A strategy profile  $\sigma = (\sigma^i)_{i \in I}$  is *stationary* if for every player  $i \in I$  the strategy  $\sigma^i$  is stationary. Denote by  $\Sigma_{\text{stat}} = \times_{i \in I} \Sigma_{\text{stat}}^i$  the set of stationary strategy profiles.

In Section 4.1 we will make use of the concept of *general strategy* (see Mertens, Sorin, and Zamir, 2015), which is a probability distribution over behavior strategies. By Kuhn's Theorem (Kuhn, 1956) every general strategy is equivalent to a behavior strategy.

Every initial state  $s_0 \in S$  and every strategy profile  $\sigma = (\sigma^i)_{i \in I} \in \Sigma$  induce a probability distribution  $\mathbf{P}_{s_0, \sigma}$  over the set of plays  $H^\infty$ . Denote the corresponding expectation operator by  $\mathbf{E}_{s_0, \sigma}$ .

For every discount factor  $\lambda \in [0, 1)$ , every initial state  $s_0$ , and every strategy profile  $\sigma$ , the  $\lambda$ -discounted payoff is

$$\gamma_\lambda^i(s_0; \sigma) := \mathbf{E}_{s_0, \sigma} \left[ (1 - \lambda) \sum_{n=0}^{\infty} \lambda^n u^i(s_n, a_n) \right]. \quad (1)$$

For every integer  $N \geq 0$ , every initial state  $s_0$ , and every strategy profile  $\sigma$ , the  $N$ -stage payoff is

$$\gamma_N^i(s_0; \sigma) := \frac{1}{N} \mathbf{E}_{s_0, \sigma} \left[ \sum_{n=0}^{N-1} u^i(s_n, a_n) \right].$$

### 3 The Modified Games: Definition and Summary of Results

The main subject of this paper is the modified game, which is an auxiliary normal-form game that corresponds to a given stochastic game  $\Gamma = (S, I, (A^i(s))_{s \in S, i \in I}, (u^i)_{i \in I}, q)$ . In this normal-form game, the set of players is  $I$  and the set of strategies of each

player  $i \in I$  is  $\Sigma^i$ , as in the stochastic game. The payoff function of the modified game depends on four elements: (a) an initial state  $s_0 \in S$ , (b) a discount factor  $\lambda \in [0, 1)$ , (c) a collection  $\vec{\mathcal{D}} = (\mathcal{D}^i)_{i \in I}$  of partitions of the set of states, one partition for each player, and (d) a collection of vectors  $\vec{c} = (c^i)_{i \in I}$ , where  $c^i = (c^i(D))_{D \in \mathcal{D}^i} \in \mathbb{R}^{\mathcal{D}^i}$  for each player  $i \in I$ . In the next subsection we define the payoff function of the modified game.

### 3.1 Definition of the modified game

**Definition 3.1** *For every strategy profile  $\sigma$ , every discount factor  $\lambda \in [0, 1)$ , every initial state  $s_0 \in S$ , every state  $s \in S$ , and every action profile  $a \in A(s)$ , the expected  $\lambda$ -discounted time the play spends in state  $s$  and the players play the action profile  $a$  is*

$$t_\lambda(s_0, \sigma; s, a) := \mathbf{E}_{s_0, \sigma} \left[ (1 - \lambda) \sum_{n=0}^{\infty} \lambda^n \mathbf{1}_{\{s_n = s, a_n = a\}} \right].$$

We denote by  $t_\lambda(s_0, \sigma) := (t_\lambda(s_0, \sigma; s, a))_{s \in S, a \in A(s)}$  the state-action discounted time vector.

Fix an initial state  $s_0 \in S$ , a discount factor  $\lambda \in [0, 1)$ , and a strategy profile  $\sigma \in \Sigma$ . For every set of states  $D \subseteq S$ , the  $\lambda$ -discounted time that the play spends in  $D$  is

$$t_\lambda(s_0, \sigma; D) := \mathbf{E}_{s_0, \sigma} \left[ (1 - \lambda) \sum_{n=0}^{\infty} \lambda^n \mathbf{1}_{\{s_n \in D\}} \right] = \sum_{s \in D, a \in A(s)} t_\lambda(s_0, \sigma; s, a).$$

Let  $i \in I$  be a player. The  $\lambda$ -discounted (unnormalized) payoff that player  $i$  receives while the play is in  $D$  is

$$\begin{aligned} U_\lambda^i(s_0, \sigma; D) &:= \mathbf{E}_{s_0, \sigma} \left[ (1 - \lambda) \sum_{n=0}^{\infty} \lambda^n u^i(s_n, a_n) \mathbf{1}_{\{s_n \in D\}} \right] \\ &= \sum_{s \in D, a \in A(s)} t_\lambda(s_0, \sigma; s, a) u^i(s, a). \end{aligned} \tag{2}$$

The quantity  $\frac{U_\lambda^i(s_0, \sigma; D)}{t_\lambda(s_0, \sigma; D)}$  is the normalized  $\lambda$ -discounted payoff of player  $i$  during visits to  $D$ .

Given a partition  $\mathcal{D}^i$  of the set of states and a vector  $c^i = (c^i(D))_{D \in \mathcal{D}^i} \in \mathbb{R}^{\mathcal{D}^i}$  define a new payoff function  $\hat{\gamma}_\lambda^i(s_0, \cdot; \mathcal{D}^i, c^i) : \Sigma \rightarrow \mathbb{R}$  over the set of strategy profiles by

$$\begin{aligned} \hat{\gamma}_\lambda^i(s_0, \sigma; \mathcal{D}^i, c^i) &:= \sum_{D \in \mathcal{D}^i} \min \{U_\lambda^i(s_0, \sigma; D), t_\lambda(s_0, \sigma; D) \cdot c^i(D)\} \\ &= \sum_{D \in \mathcal{D}^i} t_\lambda(s_0, \sigma; D) \min \left\{ \frac{U_\lambda^i(s_0, \sigma; D)}{t_\lambda(s_0, \sigma; D)}, c^i(D) \right\}, \quad \forall \sigma \in \Sigma, \end{aligned} \quad (3)$$

where by convention  $\frac{0}{0} = 1$ . Thus, to calculate the payoff  $\hat{\gamma}_\lambda^i(s_0, \sigma; \mathcal{D}^i, c^i)$ , we take the normalized  $\lambda$ -discounted payoff during the visits to each element  $D$  of  $\mathcal{D}^i$ , and, if this quantity is higher than  $c^i(D)$ , we lower it to  $c^i(D)$ . We then sum up the quantities that we obtained for all elements of  $D$ , after multiplying each one by the  $\lambda$ -discounted time the play spends in  $D$ . Accordingly, with the new payoff function  $c^i(D)$  is the maximal amount that player  $i$  can receive during the visits to  $D$ .

For every partition  $\mathcal{D}^i$  of the set of states we have

$$\gamma_\lambda^i(s_0, \sigma) = \sum_{D \in \mathcal{D}^i} U_\lambda^i(s_0, \sigma; D), \quad \forall \lambda \in [0, 1), \forall s_0 \in S, \forall i \in I, \forall \sigma \in \Sigma. \quad (4)$$

Consequently, by Eq. (3)

$$\hat{\gamma}_\lambda^i(s_0, \sigma; \mathcal{D}^i, c^i) \leq \gamma_\lambda^i(s_0, \sigma), \quad \forall \lambda \in [0, 1), \forall s_0 \in S, \forall i \in I, \forall \sigma \in \Sigma. \quad (5)$$

Since payoffs are bounded by 1, if  $c^i(D) \geq 1$  for every element  $D \in \mathcal{D}^i$ , then there is an equality in Eq. (5).

**Definition 3.2** Let  $\lambda \in [0, 1)$  be a discount factor, let  $s_0 \in S$  be a state, let  $\vec{\mathcal{D}} = (\mathcal{D}^i)_{i \in I}$  be a collection of partitions of the set of states, and let  $\vec{c} = (c^i)_{i \in I} \in \times_{i \in I} \mathbb{R}^{\mathcal{D}^i}$ . The modified games  $\hat{\Gamma}_\lambda(s_0; \vec{\mathcal{D}}, \vec{c})$  is the normal-form game  $(I, \Sigma, (\hat{\gamma}_\lambda^i(s_0, \cdot; \mathcal{D}^i, c^i))_{i \in I})$ .



**Remark 3.3 (Alternative definitions of the modified game)** *The minimum in Eq. (3) that defines the payoff in the modified game is between the discounted payoff during the visits to an element  $D$  of  $\mathcal{D}^i$  and the normalized cut-off. Several alternative definitions come to mind:*

1. *The minimum could have been taken in each stage separately, instead of dividing the stages according to the element of  $\mathcal{D}^i$  that contains the current state.*
2. *The minimum could have been taken for each play path separately.*
3. *The minimum could have been taken for each visit to an element of  $\mathcal{D}^i$  separately, instead of aggregating all visits to the same element.*

*Each alternative definition will satisfy some of the results that we prove in the sequel, but not all of them.*

**Remark 3.4 (The modified game in absorbing games)** *When specialized to absorbing games and the partition  $\mathcal{D}^i$  that contains only singletons, for every player  $i \in I$ , the modified game defined by Solan (1999) coincides with alternative 1 in Remark 3.3. Another variant of the modified game for absorbing games was presented in Solan and Vohra (2002). Unlike Solan (1999) and Solan and Vohra (2002), where the modified game was an absorbing game, just like the original game, here the modified game is a normal-form game.*

**Remark 3.5 (On the discount factor)** *Definition 3.2 assumes that all players share the same discount factor. Our results continue to hold if each player has a different discount factor.*

**Remark 3.6 (The modified game of zero-sum games)** *It is worth noting that when the original stochastic game is a two-player zero-sum game, the modified game is no longer a zero-sum game.*

### 3.2 Summary of Results

Since the paper is long and presents many results, some on the modified game and some that are used in the proofs of the main results, we end this section by providing a list of the main results concerning the modified game.

- The modified game  $\widehat{\Gamma}_\lambda(s_0; \vec{\mathcal{D}}, \vec{c})$  admits an equilibrium. Moreover, there is an equilibrium in which all players use stationary strategies (Theorem 4.1).
- The function that assigns to each discount factor  $\lambda$  the set of equilibria in stationary strategies of the modified game  $\widehat{\Gamma}_\lambda(s_0; \vec{\mathcal{D}}, \vec{c})$  is semi-algebraic (Theorem 4.11).
- Denote by  $\underline{v}_\lambda^i(s)$  the  $\lambda$ -discounted max-min value of player  $i$  in the stochastic game when the initial state is  $s$ . If the partition  $\mathcal{D}^i$  satisfies a certain property (see Definition 6.2), and if  $c^i(D) \geq \lim_{\lambda \rightarrow 1} \underline{v}_\lambda^i(s)$  for every element  $D \in \mathcal{D}^i$  and every state  $s \in D$ , then the limit of the max-min value of the modified game as the discount factor goes to 1 exists and is equal to  $\lim_{\lambda \rightarrow 1} \underline{v}_\lambda^i(s_0)$  (Theorem 6.8).
- Denote by  $\overline{v}_\lambda^i(s)$  the  $\lambda$ -discounted min-max value of player  $i$  in the stochastic game when the initial state is  $s$ . If the partition  $\mathcal{D}^i$  satisfies a certain property (see Definition 7.2), and if  $c^i(D) \geq \lim_{\lambda \rightarrow 1} \overline{v}_\lambda^i(s)$  for every element  $D \in \mathcal{D}^i$  and every state  $s \in D$ , then the limit of the min-max value of the modified game as the discount factor goes to 1 exists and is equal to  $\lim_{\lambda \rightarrow 1} \overline{v}_\lambda^i(s_0)$  (Theorem 7.3).
- The min-max value in stationary strategies of the modified game is at least the min-max value of the modified game and at most the discounted min-max value in the original game (Theorem 8.1).
- The max-min value in stationary strategies of the modified game is at most the max-min value of the modified game (Theorem 9.3), and the difference may not vanish as the discount factor goes to 1 (Example 9.4).

## 4 The Modified Stochastic Game: Equilibrium

Using standard arguments one can show that the modified game admits an equilibrium. Indeed, the space of pure strategies is compact in the product topology, hence the space of mixed strategies is compact in the weak-\* topology. Moreover, the modified payoff function, defined on the space of profiles of mixed strategies, is continuous in the weak-\* topology and concave in each of its coordinates. By, e.g., Schauder's fixed point theorem, an equilibrium exists.

The payoff function in the modified game is neither multilinear nor quasiconcave when restricted to stationary strategies, hence it is not clear that the game admits an equilibrium in stationary strategies. Nevertheless we will prove in this section that the modified game admits an equilibrium in stationary strategies.

**Theorem 4.1** *For every initial state  $s_0 \in S$ , every discount factor  $\lambda \in [0, 1)$ , every collection  $\vec{\mathcal{D}} = (\mathcal{D}^i)_{i \in I}$  of partitions of the set of states, and every vector of cutoffs  $\vec{c} = (c^i)_{i \in I} \in \times_{i \in I} \mathbb{R}^{\mathcal{D}^i}$ , the modified game  $\hat{\Gamma}_\lambda(s_0; \vec{\mathcal{D}}, \vec{c})$  admits an equilibrium in stationary strategies.*

When particularized to absorbing games, Theorem 4.1 coincides with Step 1 in the proof of Theorem 4.5 in Solan (1999). In Section 4.1 we present some technical tools that are needed to prove Theorem 4.1. The proof of the theorem together with additional results appear in Section 4.2.

### 4.1 A Result on Stationary Strategies in MDP's

Below we prove the existence of an equilibrium in the modified game by using a fixed point theorem. To this end we will consider the best response of one player to fixed strategies of the other players. When fixing the stationary strategies of all players but one, the game is reduced to a Markov decision problem. The best-response set-valued function in the modified game does not have convex values, hence we will need a new tool to prove the existence of a fixed point, which we describe in this section.

A *Markov decision process* (MDP) is a stochastic game with a single player. For notational convenience we denote the unique player by  $i$ .

We say that the two strategies  $\sigma^i$  and  $\sigma'^i$  are  $\lambda$ -*equivalent* at the initial state  $s_0$  if the two strategies induce the same state-action discounted time vector, that is,  $t_\lambda(s_0, \sigma^i) = t_\lambda(s_0, \sigma'^i)$ . As the following result states, for every strategy there exists a  $\lambda$ -discounted equivalent stationary strategy at any given initial state.

**Theorem 4.2 (Altman, 1999, Corollary 10.1)** *For every discount factor  $\lambda \in [0, 1)$ , every initial state  $s_0 \in S$ , and every strategy  $\sigma^i \in \Sigma^i$  there exists a stationary strategy  $x^i \in \Sigma_{stat}^i$  that is  $\lambda$ -equivalent to  $\sigma^i$  at  $s_0$ .*

The following result states that the stationary strategy  $x^i$  that is  $\lambda$ -equivalent to  $\sigma^i$  at  $s_0$  is unique, up to its definition in states that are never visited.

**Lemma 4.3** *Let  $\lambda \in [0, 1)$  be a discount factor, let  $s_0 \in S$  be a state, let  $x^i$  and  $x'^i$  be two stationary strategies in  $\Sigma_{stat}^i$  that are  $\lambda$ -discounted equivalent at  $s_0$ , let  $s \in S$ , and let  $a \in A^i(s)$ . If  $t_\lambda(s_0, x^i; \{s\}) > 0$  then  $x^i(s; a) = x'^i(s; a)$ .*

**Proof.** The stationarity of  $x^i$  implies that

$$x^i(s; a) = \frac{t_\lambda(s_0, x^i; s, a)}{\sum_{a' \in A^i(s)} t_\lambda(s_0, x^i; s, a')} = \frac{t_\lambda(s_0, x^i; s, a)}{t_\lambda(s_0, x^i; \{s\})},$$

whenever the denominator is positive, and a similar equality holds for  $x'^i$ . Since  $x^i$  and  $x'^i$  are  $\lambda$ -equivalent at  $s_0$ , either both denominators (for  $x^i$  and for  $x'^i$ ) are positive, or both are equal to 0. The result follows. ■

We will need below the following observation that strengthens Theorem 4.2 and that is stated without a proof. It says that if  $x^i$  is a stationary strategy that is  $\lambda$ -discounted equivalent to  $\sigma^i$ , then the probability that  $x^i$  plays the action  $a$  at state  $s$  cannot exceed the maximal probability that  $\sigma^i$  plays  $a$  in some visit to  $s$ , and cannot fall below the minimal probability that  $\sigma^i$  plays  $a$  in some visit to  $s$ .

To formally state the result, for every state  $s \in S$  we denote by  $H_s$  the set of all finite histories that end at  $s$ :

$$H_s := \{(s_0, a_0, \dots, s_n) \in H : s_n = s\}.$$

**Lemma 4.4** *Let  $\lambda \in [0, 1)$  be a discount factor, let  $s_0 \in S$  be a state, and let  $\sigma^i \in \Sigma^i$  be a strategy. Let  $x^i \in \Sigma_{stat}^i$  be a stationary strategy that is  $\lambda$ -equivalent to  $\sigma^i$  at  $s_0$ . Then*

$$\inf_{h \in H_s} \sigma^i(h; a) \leq x^i(s; a) \leq \sup_{h \in H_s} \sigma^i(h; a).$$

It is well known that the set of stationary optimal strategies in an MDP is convex. The following lemma implies in particular a weaker version of this result, namely that the set of optimal stationary strategies is monovex. Below we will use the lemma for a modified game, in which the set of stationary optimal strategies is not convex.

**Lemma 4.5** *Let  $s_0 \in S$  be a state and let  $x^i$  and  $x'^i$  be two stationary strategies in  $\Sigma_{stat}^i$ . Let  $\sigma_\alpha^i := [\alpha(x^i), (1 - \alpha)(x'^i)]$  be the general strategy that follows  $x^i$  with probability  $\alpha$  and  $x'^i$  with probability  $1 - \alpha$ . For every  $\alpha \in [0, 1]$  let  $x_\alpha^i$  be a stationary strategy that is  $\lambda$ -equivalent to the strategy  $\sigma_\alpha^i$  at the initial state  $s_0$  (see Proposition 4.2), so that*

$$t_\lambda(s_0, x_\alpha^i) = \alpha t_\lambda(s_0, x^i) + (1 - \alpha) t_\lambda(s_0, x'^i).$$

*Then the function  $\alpha \mapsto x_\alpha^i(s; a)$  is monotone for every state  $s \in S$  for which  $\sum_{a' \in A^i(s)} t_\lambda(s_0, x^i; s, a') + \sum_{a' \in A^i(s)} t_\lambda(s_0, x'^i; s, a') > 0$  and for every action  $a \in A^i(s)$ .*

**Proof.** Note that if  $\alpha = \beta\alpha' + (1 - \beta)\alpha''$  then  $x_\alpha^i$  is  $\lambda$ -equivalent at  $s_0$  to the general strategy  $[\beta(x_{\alpha'}^i), (1 - \beta)(x_{\alpha''}^i)]$ . Together with Lemma 4.4 this implies that the function  $\alpha \mapsto x_\alpha^i(s; a)$  is monotone for every state  $s \in S$  for which  $t_\lambda(s_0, x^i; \{s\}) + t_\lambda(s_0, x'^i; \{s\}) > 0$  and for every action  $a \in A^i(s)$ . The continuity of this function follows from Lemma 4.3 and from the continuity of the function  $x^i \mapsto t_\lambda(s_0, x^i)$ . ■

As a corollary we deduce the following.

**Corollary 4.6** *Let  $g : \Sigma^i \rightarrow \mathbb{R}$  be a real-valued function that satisfies the following conditions:*

1. The function  $g$  depends on its parameter only through its state-action discounted time vector: there are  $s_0 \in S$ ,  $\lambda \in (0, 1)$ , and a continuous function  $f : \Delta(S \times A) \rightarrow \mathbb{R}$  such that  $g(s_0, \sigma^i) = f(t_\lambda(s_0, \sigma^i))$  for every  $\sigma^i \in \Sigma^i$ .
2. The function  $g$  is quasiconcave: For every  $\sigma^i, \sigma'^i \in \Sigma^i$  and every  $\beta \in [0, 1]$  we have

$$g([\beta(\sigma^i), (1 - \beta)(\sigma'^i)]) \geq \beta g(\sigma^i) + (1 - \beta)g(\sigma'^i).$$

Then the set  $\operatorname{argmax}_{\sigma^i \in \Sigma^i} g(\sigma^i)$  of maximizers of  $g$  is a closed monovex set.

**Proof.** The set  $\operatorname{argmax}_{\sigma^i \in \Sigma^i} g(\sigma^i)$  is closed since the function  $f$  is continuous. The set is monovex by Lemma 4.5. ■

## 4.2 Stationary Equilibria in the Modified Game

Because for every set of states  $D \subseteq S$  the functions  $t_\lambda(s_0, \cdot; D)$  and  $U_\lambda^i(s_0, \cdot; D)$  are continuous over  $\Sigma$  and because the minimum of two continuous function is a continuous function, it follows that the payoff functions in the modified game is continuous over the space of strategy profiles  $\Sigma$ . This observation is summarized by the following lemma.

**Lemma 4.7** *For every  $\lambda \in [0, 1)$ , every initial state  $s_0 \in S$ , every player  $i \in I$ , every partition  $\mathcal{D}^i$  of the set of states, and every vector  $c^i \in \mathbb{R}^{\mathcal{D}^i}$ , the function  $\sigma \mapsto \hat{\gamma}_\lambda^i(s_0, \sigma; \mathcal{D}^i, c^i)$  is continuous.*

We will now show that in the modified game, when the other players play a stationary strategy, player  $i$  has a stationary best response.

**Lemma 4.8** *For every  $\lambda \in [0, 1)$ , every initial state  $s_0 \in S$ , every player  $i \in I$ , every partition  $\mathcal{D}^i$  of the set of states, every vector  $c^i \in \mathbb{R}^{\mathcal{D}^i}$ , and every stationary strategy profile  $x^{-i} \in \Sigma_{stat}^{-i}$  of the other players, there is a stationary strategy  $x^i \in \Sigma_{stat}^i$  that maximizes player  $i$ 's payoff in the modified game:*

$$\hat{\gamma}_\lambda^i(s_0, x^i, x^{-i}; \mathcal{D}^i, c^i) = \max_{\sigma^i \in \Sigma^i} \hat{\gamma}_\lambda^i(s_0, \sigma^i, x^{-i}; \mathcal{D}^i, c^i). \quad (6)$$

**Proof.** Fixing a stationary strategy profile  $x^{-i}$  of the other players, the decision problem of player  $i$  in the modified game becomes a Markov decision problem, albeit with the modified payoff function.

The set of all strategies of player  $i$  is compact and the payoff function  $\sigma^i \mapsto \hat{\gamma}_\lambda^i(s_0, \sigma^i, x^{-i}; \mathcal{D}^i, c^i)$  is continuous (Lemma 4.7), hence the maximum on the right-hand side of Eq. (6) is attained by some strategy  $\sigma^i \in \Sigma^i$ . By Proposition 4.2, at the initial state  $s_0$  the strategy  $\sigma^i$  induces the same state-action discounted time vector as some stationary strategy  $x^i \in \Sigma_{\text{stat}}^i$ . By Eq. (2) the payoff  $\hat{\gamma}_\lambda^i(s_0, \sigma^i, x^{-i}; \mathcal{D}^i, c^i)$  depends on  $\sigma^i$  only through the state-action  $\lambda$ -discounted time vector, and therefore  $\hat{\gamma}_\lambda^i(s_0, \sigma^i, x^{-i}; \mathcal{D}^i, c^i) = \hat{\gamma}_\lambda^i(s_0, x^i, x^{-i}; \mathcal{D}^i, c^i)$ , and the claim follows. ■

In the proof of Theorem 4.1 we will use sets in a Euclidean space that satisfy a condition that resembles convexity; specifically, sets that satisfy the following condition: for every two points in the set there is a continuous path in the set that connects the two points and that is monotone in each coordinate. Such sets, which were introduced in Buhovsky, Solan, and Solan (2016), are called *monovex*.

**Definition 4.9** A set  $X \subseteq \mathbb{R}^d$  is *monovex* if for every  $x, y \in X$  there is a continuous path  $f : [0, 1] \rightarrow X$  that satisfies the following properties:

1.  $f(0) = x$  and  $f(1) = y$ .
2.  $f_i(t) : [0, 1] \rightarrow \mathbb{R}$  is a monotone function (nondecreasing or nonincreasing) for every  $i \in \{1, 2, \dots, d\}$ .

The main property of monovex sets that we will need is that every closed monovex set is contractible (see Buhovsky, Solan, and Solan, 2016). We are now ready to prove Theorem 4.1.

**Proof of Theorem 4.1.** Define a set-valued function  $F : X \rightarrow X$  as follows. For every player  $i \in I$  and every stationary strategy profile  $x \in X$ ,

$$F^i(x) := \left\{ y^i \in \Sigma_{\text{stat}}^i : \hat{\gamma}_\lambda^i(s_0, y^i, x^{-i}; \mathcal{D}^i, c^i) = \max_{\sigma^i \in \Sigma^i} \hat{\gamma}_\lambda^i(s_0, \sigma^i, x^{-i}; \mathcal{D}^i, c^i) \right\}.$$

By Lemma 4.8 the set  $F^i(x)$  is nonempty for every player  $i \in I$  and every stationary strategy profile  $x \in \Sigma_{\text{stat}}$ . By Corollary 4.6 this set is monovex, and therefore by Buhovsky, Solan, and Solan (2016) it is contractible. Lemma 4.7 implies that the graph of  $F$  is closed. By the Eilenberg and Montgomery (1946) fixed point theorem the set-valued function  $F$  has a fixed point, which is a stationary equilibrium in the modified game. ■

As the following example shows, the equilibrium in stationary strategies that is guaranteed to exist by Theorem 4.1 may depend on the initial state  $s_0$ , even if the set of players includes a single player.

**Example 4.10** *Consider the following stochastic game with four states, which depends on two parameters,  $y \in \mathbb{R}$  and  $p \in [0, 1]$ , and is graphically given in Figure 1; in this figure, the payoff appears on the left-hand side of each entry, while the transition appears on the right-hand side.*

- *There is a single player:  $I = \{i\}$ .*
- *There are four states  $S = \{s^0, s^1, s^2, s^3\}$ . States  $s^2$  and  $s^3$  are absorbing and yield payoffs 2 and 3, respectively.*
- *In state  $s^0$  the player has a single action. His payoff is  $y$  and the play moves to state  $s^1$ .*
- *In state  $s^1$  the player has two actions,  $T$  and  $B$ . The payoff in this state and the transition are given in Figure 1.*

		$T$	<table><tr><td>0</td><td><math>[(1-p)(s^1), p(s^2)]</math></td></tr></table>		0	$[(1-p)(s^1), p(s^2)]$								
0	$[(1-p)(s^1), p(s^2)]$													
<table><tr><td><math>y</math></td><td><math>s^1</math></td></tr></table>		$y$	$s^1$	$B$	<table><tr><td>-1</td><td><math>[(1-p)(s^1), p(s^3)]</math></td></tr></table>		-1	$[(1-p)(s^1), p(s^3)]$	<table><tr><td>2</td><td><math>s^2</math></td></tr></table>	2	$s^2$	<table><tr><td>3</td><td><math>s^3</math></td></tr></table>	3	$s^3$
$y$	$s^1$													
-1	$[(1-p)(s^1), p(s^3)]$													
2	$s^2$													
3	$s^3$													
$s^0$		$s^1$		$s^2$		$s^3$								

Figure 1: The game in Example 4.10.



Let  $\mathcal{D}^i = \{\{s^0, s^1\}, \{s^2\}, \{s^3\}\}$ , and  $c^i = (c^i(D))_{D \in \mathcal{D}^i}$  be the vector

$$c^i(\{s^0, s^1\}) = 0, \quad c^i(\{s^2\}) = 2, \quad c^i(\{s^3\}) = 3.$$

Thus, the payoff may be lowered only in the set  $\{s^0, s^1\}$ . We will show that for a proper choice of the parameters  $\lambda$ ,  $p$ , and  $y$ , the optimal strategy in the modified game  $\widehat{\Gamma}_\lambda(s_0; \mathcal{D}^i, c^i)$  depends on the initial state: when the initial state is  $s^1$  the optimal strategy consists of playing  $T$  in state  $s^1$ , while when the initial state is  $s^0$  the optimal strategy consists of playing  $B$  in state  $s^1$ .

The intuition for this result is as follows: When the initial state is  $s^1$  the payoff while the play is in  $\{s^0, s^1\}$  is nonpositive, hence the minimum with  $c^i(\{s^0, s^1\})$  has no effect on the payoff in the modified game, and the modified game is reduced to a standard Markov decision problem. The parameters  $p$  and  $\lambda$  will be chosen so that the optimal strategy is to play  $T$  in this case. Suppose now that the initial state is  $s^0$ . Since  $y$  is positive, the player is better off choosing  $B$  at state  $s^1$  as long as his total discounted payoff before absorption is nonnegative. This increases the probability to be absorbed in the better absorbing state  $s^3$ , while not affecting the modified payoff that he receives for the stages in which the play visits the set  $\{s^0, s^1\}$ .

We turn to the formal calculations. As mentioned before, when the initial state is  $s^1$  the modified game is a standard Markov decision problem, hence there is an optimal strategy which is pure and stationary. Since at state  $s^1$  the play is absorbed at every stage with probability  $p$ , the total discounted weight of the absorbing state on the payoff is

$$\lambda p + (1-p)\lambda^2 p + (1-p)^2 \lambda^3 p + \cdots = \frac{\lambda p}{1 - \lambda(1-p)}.$$

It follows that when the initial state is  $s^1$ , the stationary strategy  $T$  is the unique optimal strategy as soon as

$$2 \frac{\lambda p}{1 - \lambda(1-p)} > -1 \left( 1 - \frac{\lambda p}{1 - \lambda(1-p)} \right) + 3 \frac{\lambda p}{1 - \lambda(1-p)},$$

which solves to

$$1 > \lambda(1+p).$$

Suppose now that the initial state is  $s^0$ . We choose  $y$  so that under the stationary strategy that plays  $B$  in  $s^0$ , the discounted payoff when the play is in  $\{s^0, s^1\}$  is 0. That is,

$$(1 - \lambda)y - \lambda \left( 1 - \frac{\lambda p}{1 - \lambda(1 - p)} \right) = 0,$$

which solves to

$$y = \frac{\lambda}{1 - \lambda(1 - p)}. \quad (7)$$

This implies that when  $y$ ,  $\lambda$ , and  $p$  satisfy Eq. (7), under every strategy, the modified payoff in the set  $\{s^0, s^1\}$  is 0, hence the optimal strategy is the one that maximizes the expected absorbing payoff, which is the stationary strategy that plays  $B$  at state  $s^1$ .

We end this section by extending the semi-algebraic property of the set of discounted equilibria of stochastic games to the modified game. Denote by  $E_\lambda(s_0; \vec{\mathcal{D}}, \vec{c})$  the set of all stationary equilibria of the game  $\hat{\Gamma}_\lambda(s_0; \vec{\mathcal{D}}, \vec{c})$ . Because for every initial state  $s_0 \in S$ , every set of states  $D \subseteq S$ , and every player  $i \in I$  the functions  $(\lambda, x) \mapsto t_\lambda(s_0, x; D)$  and  $(\lambda, x) \mapsto U_\lambda^i(s_0, x; D)$  are semi-algebraic, where  $x$  ranges over all stationary strategy profiles, we obtain the following result.

**Theorem 4.11** *For every initial state  $s_0 \in S$ , every collection  $\vec{\mathcal{D}} = (\mathcal{D}^i)_{i \in I}$  of partitions of the set of states, and every collection of cutoffs  $\vec{c} = (c^i)_{i \in I} \in \times_{i \in I} \mathbb{R}^{\mathcal{D}^i}$  the set-valued function  $\lambda \mapsto E_\lambda(s_0; \vec{\mathcal{D}}, \vec{c})$  is semi-algebraic.*

## 5 The Max-Min Value in Stochastic Games

Our next goal is to study the max-min value in the modified game. To this end we recall the concept of max-min value in the original stochastic game and a result of Neyman (2003) on the uniform max-min value in stochastic game.

The  $\lambda$ -discounted max-min value of player  $i$  at the initial state  $s_0$  in the original game is given by

$$\underline{v}_\lambda^i(s_0) := \max_{\sigma^i \in \Sigma^i} \min_{\sigma^{-i} \in \Sigma^{-i}} \gamma_\lambda^i(s_0; \sigma^i, \sigma^{-i}). \quad (8)$$

This is the maximal amount that player  $i$  can guarantee if the other players get to know his strategy and try to minimize his payoff. Because the  $\lambda$ -discounted payoff (see Eq. (1)) is a continuous function of the strategies of the players, the maximum and minima in Eq. (8) are attained. It is well known that the maximum and minima in Eq. (8) are attained by stationary strategies. Moreover, the function  $\lambda \mapsto \underline{v}_\lambda^i(s_0)$  is semi-algebraic (see Bewley and Kohlberg (1976) and Neyman (2003)), and therefore the limit

$$\underline{v}_1^i(s_0) := \lim_{\lambda \rightarrow 1} \underline{v}_\lambda^i(s_0)$$

exists for every player  $i \in I$  and every initial state  $s_0 \in S$ . The quantity  $\underline{v}_1^i(s_0)$  is called the *uniform max-min value* of player  $i$  at the initial state  $s_0$ .

For every two bounded stopping times  $\tau < \tau'$ , the expected average payoff between stages  $\tau$  and  $\tau'$  is

$$\gamma^i(s_0, \sigma; \tau, \tau') := \frac{\mathbf{E}_{s_0, \sigma}[\sum_{n=\tau}^{\tau'} r(s_n, a_n) \mid \mathcal{H}(\tau)]}{\mathbf{E}_{s_0, \sigma}[\tau - \tau' \mid \mathcal{H}(\tau)]}.$$

Note that  $\gamma^i(s_0, \sigma; \tau, \tau')$  is a random variable that is measurable according to the information at stage  $\tau$ .

**Definition 5.1** *Let  $\varepsilon > 0$  and let  $i \in I$  be a player. A strategy  $\sigma^i$  is a uniform  $\varepsilon$ -max-min strategy of player  $i$  if there exist  $\lambda_0 \in [0, 1)$  and  $N_0 \in \mathbb{N}$  such that the following holds for every initial state  $s_0 \in S$ , every strategy profile  $\sigma^{-i} \in \Sigma^{-i}$  of the other players, every discount factor  $\lambda \in (\lambda_0, 1)$ , and every bounded stopping time  $\tau$  that satisfies  $\mathbf{E}_{s_0, \sigma^i, \sigma^{-i}}[\tau] \geq N_0$ :*

$$\gamma_\lambda^i(s_0, \sigma^i, \sigma^{-i}) \geq \underline{v}_1^i(s_0) - \varepsilon, \quad (9)$$

$$\gamma^i(s_0, \sigma^i, \sigma^{-i}; 0, \tau) \geq \underline{v}_1^i(s_0) - \varepsilon. \quad (10)$$

**Remark 5.2 (On the definition of uniform  $\varepsilon$ -max-min strategies)** *The standard definition of uniform  $\varepsilon$ -max-min strategies requires Eq. (10) only for constant stopping times  $\tau$ . We will need in the sequel the stronger version that we presented here. Sometimes a uniform  $\varepsilon$ -max-min strategy is required to satisfy that the expected long-run average payoff is at least  $\underline{v}_1^i(s_0) - \varepsilon$  for every strategy profile of the*

other players (see Mertens and Neyman (1981) and Neyman (2003)). Our results are not affected by adding this requirement.

For every player  $i \in I$ , every strategy  $\sigma^i \in \Sigma^i$ , and every finite history  $h_n \in H$ , denote by  $\sigma_{h_n}^i$  the strategy profile  $\sigma$  conditioned on the history  $h_n$ , that is,

$$\sigma_{h_n}^i(\hat{h}_m) := \sigma^i(s_0, a_0, \dots, s_{n-1}, a_{n-1}, \hat{s}_0, \hat{a}_0, \hat{s}_1, \hat{a}_1, \dots, \hat{s}_m), \quad \forall \hat{h}_m = (\hat{s}_0, \hat{a}_0, \dots, \hat{s}_m) \in H.$$

**Definition 5.3** *Let  $\varepsilon > 0$  and let  $i \in I$  be a player. A strategy  $\sigma^i$  of player  $i$  is  $\varepsilon$ -max-min preserving if for every strategy profile  $\sigma^{-i} \in \Sigma^{-i}$ , every play  $h \in H^\infty$ , and every two bounded stopping times  $\tau < \tau'$  we have*

$$\mathbf{E}_{s_\tau, \sigma_{h_\tau}^i, \sigma_{h_\tau}^{-i}}[\underline{v}_1^i(s_{\tau'})] \geq \underline{v}_1^i(s_\tau) - \varepsilon. \quad (11)$$

A uniform  $\varepsilon$ -max-min strategy need not be  $\varepsilon$ -max-min preserving, and an  $\varepsilon$ -max-min preserving strategy need not be a uniform  $\varepsilon$ -max-min strategy.

Neyman (2003) adapted the construction of Mertens and Neyman (1981) and constructed for every  $\varepsilon > 0$  a uniform  $\varepsilon$ -max-min strategy that is also  $\varepsilon$ -max-min preserving.

**Theorem 5.4 (Neyman, 2003)** *For every  $\varepsilon > 0$  and every player  $i \in I$ , there is a uniform  $\varepsilon$ -max-min strategy  $\sigma^i$  that is also  $\varepsilon$ -max-min preserving.*

## 5.1 Markovian $\varepsilon$ -Max-Min Strategies

A strategy of a player is Markovian w.r.t. a partition if the mixed action played at each stage depends only on the play during the current visit to the element of the partition that contains the current state. Formally,

**Definition 5.5** *Let  $i \in I$  be a player and let  $\mathcal{D}$  be a partition of the set of states. A strategy  $\sigma^i \in \Sigma^i$  is  $\mathcal{D}$ -Markovian if  $\sigma_{h_n}^i = \sigma_{\tilde{h}_n}^i$  for every two finite histories  $h_n = (s_0, a_0, \dots, s_n)$  and  $\tilde{h}_n = (\tilde{s}_0, \tilde{a}_0, \dots, \tilde{s}_n)$  that satisfy*

- *The two histories end at the same state:  $s_n = \tilde{s}_n$ .*

- In stage  $n$  (resp. in stage  $\tilde{n}$ ) the finite history  $h_n$  (resp.  $\tilde{h}_{\tilde{n}}$ ) enters a new element of  $\mathcal{D}$ , that is,  $D(s_{n-1}) \neq D(s_n)$  and  $D(\tilde{s}_{\tilde{n}-1}) \neq D(\tilde{s}_{\tilde{n}})$ , where  $D(s)$  is the element of the partition  $\mathcal{D}$  that contains  $s$ , for every state  $s \in S$ ,

One naive way to define a  $\mathcal{D}$ -Markovian strategy from a strategy  $\sigma^i$  is the following: whenever the play enters an element of  $\mathcal{D}$ , player  $i$  forgets past play and restarts playing  $\sigma^i$ . We will denote this strategy  $\sigma_{\mathcal{D}}^i$  and define it now formally.

Given a partition  $\mathcal{D}$  of the set of states, let  $(\tau_k^{\mathcal{D}})_{k \geq 0}$  be the sequence of stopping times that indicates when the play moves from one element of the partition  $\mathcal{D}$  to another element of the partition:

$$\begin{aligned}\tau_0^{\mathcal{D}} &:= 0, \\ \tau_k^{\mathcal{D}} &:= \min \{n > \tau_{k-1}^{\mathcal{D}} : D(s_n) \neq D(s_{n-1})\}, \quad k \in \mathbb{N},\end{aligned}$$

where the minimum of an empty set is  $+\infty$ . The stages between stage  $\tau_k^{\mathcal{D}}$  and  $\tau_{k+1}^{\mathcal{D}} - 1$  are called the a  $\mathcal{D}$ -run. For every  $n \geq 0$  let  $k(\mathcal{D}; n)$  be the number of the  $\mathcal{D}$ -run that contains stage  $n$ ; that is, it is the unique nonnegative integer  $k$  that satisfies  $\tau_k^{\mathcal{D}} \leq n < \tau_{k+1}^{\mathcal{D}}$ .

Denote by  $\varphi(h_n)$  the play along  $h_n$  since the last switch of an element in  $\mathcal{D}$ :

$$\varphi(h_n) := (s_{\tau_{k(\mathcal{D},n)}^{\mathcal{D}}(h_n)}^{\mathcal{D}}, a_{\tau_{k(\mathcal{D},n)}^{\mathcal{D}}(h_n)}^{\mathcal{D}}, s_{\tau_{k(\mathcal{D},n)}^{\mathcal{D}}(h_n)+1}^{\mathcal{D}}, a_{\tau_{k(\mathcal{D},n)}^{\mathcal{D}}(h_n)+1}^{\mathcal{D}}, \dots, s_n).$$

The strategy  $\sigma_{\mathcal{D}}^i$  that is defined by

$$\sigma_{\mathcal{D}}^i(h_n) := \sigma^i(\varphi(h_n)), \quad \forall h_n \in H,$$

is  $\mathcal{D}$ -Markovian.

For every play  $h \in H^\infty$  and every partition  $\mathcal{D}$  of the set of states, denote by  $Z(h; \mathcal{D}) \in \{0, 1, 2, \dots, \infty\}$  the number of times in which the play switches between elements of  $\mathcal{D}$  along  $h$ :

$$Z(h; \mathcal{D}) := \sup\{k \geq 0 : \tau_k^{\mathcal{D}} < +\infty\}.$$

## 5.2 The Partition According to the Uniform Max-Min Value of Player $i$

It will be useful to single out the partition  $\mathcal{D}_*^i$  of the set of states  $S$  according to the uniform max-min value of player  $i$ : two states  $s, s' \in S$  are in the same element of the partition  $\mathcal{D}_*^i$  if and only if  $\underline{v}_1^i(s) = \underline{v}_1^i(s')$ . The following result implies that when player  $i$  plays an  $\varepsilon$ -max-min preserving strategy, the expected number of times that the play moves between elements of  $\mathcal{D}_*^i$  is bounded by a constant that is independent of  $\varepsilon$  and the strategy profile of the other players. Before stating the result we provide a weaker version of the concept of an  $\varepsilon$ -max-min preserving strategy.

**Definition 5.6** *Let  $i \in I$  be a player, let  $\varepsilon > 0$ , and let  $\mathcal{D}$  be a partition of the set of states. A strategy  $\sigma^i \in \Sigma^i$  is  $(\mathcal{D}, \varepsilon)$ -max-min preserving if for every initial state  $s_0 \in S$ , every strategy profile of the other players  $\sigma^{-i} \in \Sigma^{-i}$ , every play  $h \in H^\infty$ , every  $k \geq 0$ , and every bounded stopping time  $\tau_k^{\mathcal{D}} < \tau' \leq \tau_{k+1}^{\mathcal{D}}$  we have*

$$\mathbf{E}_{s_{\tau_k^{\mathcal{D}_*^i}}, \sigma_h^i, \sigma_h^{-i}} [\underline{v}_1^i(s_{\tau'})] \geq \underline{v}_1^i(s_{\tau_k^{\mathcal{D}_*^i}}) - \varepsilon.$$

Note that every  $\varepsilon$ -max-min preserving strategy is in particular  $(\mathcal{D}, \varepsilon)$ -max-min preserving, for every partition  $\mathcal{D}$  of the set of states. Moreover, if the strategy  $\sigma^i$  is  $\varepsilon$ -max-min preserving, then for every partition  $\mathcal{D}$  the strategy  $\sigma_{\mathcal{D}}^i$  is  $(\mathcal{D}, \varepsilon)$ -max-min preserving.

Denote the minimal distance between distinct uniform max-min values of some player by

$$\rho := \min\{|\underline{v}_1^i(s) - \underline{v}_1^i(s')| : s, s' \in S, i \in I, \underline{v}_1^i(s) \neq \underline{v}_1^i(s')\}.$$

**Theorem 5.7** *For every  $\delta > 0$  there is  $\varepsilon_0 > 0$  such that for every  $\varepsilon < \varepsilon_0$ , every player  $i \in I$ , every  $(\mathcal{D}_*^i, \varepsilon)$ -max-min preserving strategy  $\sigma^i \in \Sigma^i$ , and every strategy profile  $\sigma^{-i} \in \Sigma^{-i}$  there is an event  $E$  for which  $\mathbf{P}_{s_0, \sigma^i, \sigma^{-i}}(E) > 1 - \delta$  and*

$$\mathbf{E}_{s_0, \sigma^i, \sigma^{-i}}[Z(\cdot; \mathcal{D}_*^i) \cdot \mathbf{1}_E] \leq C(\rho), \quad (12)$$

where  $C(\rho)$  is some constant that is independent of  $\delta$ .

**Proof.** We first recall a general result on submartingales. Let  $\eta > 0$  and let  $(W_n)_{n \geq 0}$  be a submartingale that satisfies the following properties for every  $n \geq 0$ :

- The random variable  $W_n$  attains values in the interval  $[-1, 2]$ .
- Either  $W_{n+1} = W_n$  or  $|W_{n+1} - W_n| \geq \eta$ .

Denote by  $Z^*$  the number of times in which the process  $(W_n)$  changes its value:

$$Z^* := \#\{n \geq 0 : W_n \neq W_{n+1}\}. \quad (13)$$

Then there is a constant  $C(\eta)$  such that  $\mathbf{E}[Z^*] \leq C(\eta)$ . Moreover, the function  $\eta \mapsto C(\eta)$  can be taken to be monotonic nonincreasing. This result can be deduced by the bound on the expected number of downcrossings of a bounded submartingale, see, e.g., Billingsley, 1995, Theorem 35.4.

Let now  $\delta > 0$  and fix two real numbers  $K \geq \frac{C(\frac{\rho}{2})}{\delta}$  and  $\varepsilon_0 < \frac{\rho}{2K}$ . Define a stochastic process  $(W_n)_{n=0}^\infty$  by

$$W_n := \begin{cases} v_1^i(s_n) + \varepsilon_0 k(\mathcal{D}_*^i; n), & W_{n-1} < 1 + \frac{\rho}{2}, \\ W_{n-1} & W_{n-1} \geq 1 + \frac{\rho}{2}. \end{cases}$$

Note that  $W_n < 1 + \rho$  for every  $n \geq 0$ . Ignoring the upper bound of  $1 + \frac{\rho}{2}$  that we impose on the process  $(W_n)_{n \in \mathbb{N}}$ , this process is equal to the uniform max-min value of the current state, plus  $\varepsilon$  multiplied by the number of times in which the uniform max-min value changed along the play.

Let  $\varepsilon < \varepsilon_0$ , let  $\sigma^i$  be a  $(\mathcal{D}_*^i, \varepsilon)$ -max-min preserving strategy of player  $i$ , and let  $\sigma^{-i}$  be any strategy profile of the other players. Since  $\sigma^i$  is a  $(\mathcal{D}_*^i, \varepsilon)$ -max-min preserving strategy, the process  $(W_n)_{n=0}^\infty$  is a submartingale under  $\mathbf{P}_{s_0, \sigma^i, \sigma^{-i}}$ . Whenever the uniform max-min value of player  $i$  changes, the value of  $W_n$  changes by at least  $\rho - \varepsilon_0$ . By Eq. (13),  $\mathbf{E}_{s_0, \sigma^i, \sigma^{-i}}[Z^*] \leq C(\rho - \varepsilon_0) \leq C(\frac{\rho}{2})$ . By Markov's Inequality,  $\mathbf{P}_{s_0, \sigma^i, \sigma^{-i}}(Z^* < K) \geq 1 - \frac{C(\frac{\rho}{2})}{K} \geq 1 - \delta$ . Define the event  $E := \{Z^* < K\}$ . On this event we have  $W_n \leq 1 + K\varepsilon < 1 + \frac{\rho}{2}$  for every  $n \geq 0$ , and therefore on this event  $Z(\cdot; \mathcal{D}_*^i) = Z^*$ . Moreover,  $\mathbf{P}_{s_0, \sigma^i, \sigma^{-i}}(E) \geq 1 - \delta$ , and

$$\mathbf{E}_{s_0, \sigma^i, \sigma^{-i}}[Z(\cdot; \mathcal{D}_*^i) \cdot \mathbf{1}_E] \leq \mathbf{E}_{s_0, \sigma^i, \sigma^{-i}}[Z^*] \leq C(\frac{\rho}{2}),$$

and the result follows. ■

As a conclusion of Theorem 5.7 we deduce that in the notations of the statement of the theorem, if the strategy  $\sigma^i$  is in addition  $\mathcal{D}_*^i$ -Markovian, then the expected number of times the uniform max-min value of player  $i$  changes along the play is bounded.

**Corollary 5.8** *In the notations of Theorem 5.7, if the strategy  $\sigma^i$  is in addition  $\mathcal{D}_*^i$ -Markovian then  $\mathbf{E}_{s_0, \sigma^i, \sigma^{-i}}[Z(\cdot; \mathcal{D}_*^i)] \leq \frac{4C(\rho)}{1-\delta}$ .*

**Proof.** By Eq. (12) and Markov's inequality,  $\mathbf{P}_{s_0, \sigma^i, \sigma^{-i}}(Z(\cdot; \mathcal{D}_*^i) \cdot \mathbf{1}_E) \leq 2C(\rho) \geq \frac{1}{2}(1-\delta)$ . Denoting  $x := \mathbf{E}_{s_0, \sigma^i, \sigma^{-i}}[Z(\cdot; \mathcal{D}_*^i)]$  we deduce that  $x \leq 2C(\rho) + \frac{1}{2}(1+\delta)x$ , and the claim follows. ■

The following result<sup>2</sup> implies in particular that every player  $i$  has a  $\mathcal{D}_*^i$ -Markovian uniform  $\varepsilon$ -max-min strategy, for every  $\varepsilon > 0$ .

**Theorem 5.9** *For every  $\delta > 0$  there is  $\varepsilon_0 > 0$  such that for every  $\varepsilon < \varepsilon_0$  and every uniform  $\varepsilon$ -max-min strategy  $\sigma^i$  that is  $\varepsilon$ -max-min preserving, the strategy  $\sigma_{\mathcal{D}_*^i}^i$  is uniform  $\delta$ -max-min and  $(\mathcal{D}_*^i, \varepsilon)$ -max-min preserving.*

**Proof.** Fix  $\delta > 0$  and a player  $i \in I$ . Let  $\varepsilon$  be smaller than the quantity  $\varepsilon_0$  given in Theorem 5.7 and smaller than  $\frac{\delta\rho}{2C(\rho)}$ , where  $C(\rho)$  is the constant given in Theorem 5.7. Let  $\sigma^i$  be a uniform  $\varepsilon$ -max-min strategy that is  $\varepsilon$ -max-min preserving and fix a strategy profile of the other players  $\sigma^{-i} \in \Sigma^{-i}$ .

By construction, the strategy  $\sigma_{\mathcal{D}_*^i}^i$  is  $(\mathcal{D}_*^i, \varepsilon)$ -max-min preserving. It remains to show that the strategy  $\sigma_{\mathcal{D}_*^i}^i$  is a uniform  $\delta$ -max-min strategy.

For every two stopping times  $\tau$  and  $\tau'$  denote their minimum by

$$\tau \wedge \tau' := \min\{\tau, \tau'\}.$$

---

<sup>2</sup>Abraham Neyman advised the author that Jean-François Mertens and himself were aware of this result.



Since the strategy  $\sigma_{\mathcal{D}_*^i}^i$  is  $(\mathcal{D}_*^i, \varepsilon)$ -max-min preserving, for every  $k \geq 0$  we have

$$v(s_{\tau_k^{\mathcal{D}_*^i} \wedge \tau}) \leq \mathbf{E}_{s_0, \sigma_{\mathcal{D}_*^i}^i, \sigma^{-i}}[v(s_{\tau_{k+1}^{\mathcal{D}_*^i} \wedge \tau})] - \varepsilon \quad \text{on the event } \{\tau_k^{\mathcal{D}_*^i} \wedge \tau < \tau_{k+1}^{\mathcal{D}_*^i} \wedge \tau\}. \quad (14)$$

Using iteratively Eq. (14) over all  $k \geq 0$  we deduce that for every bounded stopping time  $\tau$

$$\begin{aligned} \underline{v}_1^i(s_0) &\leq \mathbf{E}_{s_0, \sigma_{\mathcal{D}_*^i}^i, \sigma^{-i}}[\underline{v}_1^i(s_\tau)] - \varepsilon \left( 1 + \sum_{k \geq 0} \mathbf{P}_{s_0, \sigma_{\mathcal{D}_*^i}^i, \sigma^{-i}}(\tau_k^{\mathcal{D}_*^i} \wedge \tau < \tau_{k+1}^{\mathcal{D}_*^i} \wedge \tau) \right) \\ &\leq \mathbf{E}_{s_0, \sigma_{\mathcal{D}_*^i}^i, \sigma^{-i}}[\underline{v}_1^i(s_\tau)] - \varepsilon C_2(\rho), \end{aligned}$$

where  $C_2(\rho) = \frac{4C(\rho)}{1-\delta}$  and the last inequality follows from Corollary 5.8. By averaging this inequality over all constant stopping times  $\tau \in \{0, 1, \dots, N\}$  we conclude that for every  $N \geq 0$ ,

$$\underline{v}_1^i(s_0) \leq \frac{1}{N} \sum_{n=0}^{N-1} \underline{v}_1^i(s_n) - \varepsilon C_2(\rho). \quad (15)$$

By Theorem 5.7 there is an event  $E$  for which  $\mathbf{P}_{s_0, \sigma_{\mathcal{D}_*^i}^i, \sigma^{-i}}(E) > 1 - \delta$  and

$$\mathbf{E}_{s_0, \sigma_{\mathcal{D}_*^i}^i, \sigma^{-i}}[Z(\cdot; \mathcal{D}_*^i) \cdot \mathbf{1}_E] \leq C(\rho). \quad (16)$$

By Eq. (16) and Markov's inequality we deduce that

$$\begin{aligned} \mathbf{P}_{s_0, \sigma_{\mathcal{D}_*^i}^i, \sigma^{-i}}(Z(\cdot; \mathcal{D}_*^i) < \frac{\rho}{2\varepsilon} \mid E) &\geq 1 - \frac{2\varepsilon}{\rho} \mathbf{E}_{s_0, \sigma_{\mathcal{D}_*^i}^i, \sigma^{-i}}[Z(\cdot; \mathcal{D}_*^i) \mid E] \\ &\geq 1 - \frac{2\varepsilon}{\rho} C(\rho) \geq 1 - \delta, \quad \forall \sigma^{-i} \in \Sigma^{-i}. \end{aligned} \quad (17)$$

Denote by  $F$  the event

$$F := E \cap \left\{ Z(\cdot; \mathcal{D}_*^i) < \frac{\rho}{2\varepsilon} \right\}.$$

Since  $\mathbf{P}_{s_0, \sigma_{\mathcal{D}_*^i}^i, \sigma^{-i}}(E) > 1 - \delta$  and by Eq. (17),  $\mathbf{P}_{s_0, \sigma_{\mathcal{D}_*^i}^i, \sigma^{-i}}(F) \geq 1 - 2\delta$ . Since the strategy  $\sigma_{\mathcal{D}_*^i}^i$  is  $\mathcal{D}_*^i$ -Markovian and since the strategy  $\sigma^i$  is  $\varepsilon$ -max-min preserving,

for every  $k \geq 0$  and every bounded stopping time  $\tau'$  that satisfies  $\tau_k^{\mathcal{D}_*^i} < \tau' \leq \tau_{k+1}^{\mathcal{D}_*^i}$  we have

$$\gamma^i(s_0, \sigma_{\mathcal{D}_*^i}^i, \sigma^{-i}; \tau_k^{\mathcal{D}_*^i}, \tau') \geq \underline{v}_1^i(s_{\tau_k^{\mathcal{D}_*^i}}) - \varepsilon \quad \text{on the event} \quad \left\{ \mathbf{E}_{s_0, \sigma_{\mathcal{D}_*^i}^i, \sigma^{-i}}[\tau' - \tau_k^{\mathcal{D}_*^i} \mid \mathcal{H}(\tau_k^{\mathcal{D}_*^i})] \geq N_0 \right\}, \quad (18)$$

where  $N_0$  is the constant given in Theorem 5.4.

On the event  $F$  the strategy  $\sigma^i$  restarts at most  $\frac{\rho}{2\varepsilon}$  times, and therefore on this event the expected number of stages that take part in  $\mathcal{D}_*^i$ -runs shorter than  $N_0$  is at most  $\frac{\rho N_0}{2\varepsilon}$ . Since  $\mathbf{P}_{s_0, \sigma_{\mathcal{D}_*^i}^i, \sigma^{-i}}(F) > 1 - 2\delta$ , since payoffs are bounded by 1, and by Eqs. (15) and (18) we deduce that for every stopping time  $\tau$  that satisfies  $\mathbf{E}_{s_0, \sigma_{\mathcal{D}_*^i}^i, \sigma^{-i}}[\tau] \geq \frac{\rho N_0}{2\varepsilon^2}$  we have

$$\gamma^i(s_0, \sigma_{\mathcal{D}_*^i}^i, \sigma^{-i}; 0, \tau) \geq \underline{v}_1^i(s_0) - 3\varepsilon - 4\delta,$$

and therefore for every  $\lambda$  sufficiently close to 1,

$$\gamma_\lambda^i(s_0, \sigma_{\mathcal{D}_*^i}^i, \sigma^{-i}) \geq \underline{v}_1^i(s_0) - 4\varepsilon - 4\delta.$$

The desired result follows. ■

A useful property of uniform  $\varepsilon$ -max-min  $\mathcal{D}_*^i$ -Markovian strategies is that if the expected length of the  $k$ 'th run is high, then the average expected payoff during the  $k$ 'th run is high. This observation is summarized in the following Lemma.

**Lemma 5.10** *Let  $i \in I$  be a player, let  $\varepsilon > 0$ , let  $\sigma^i \in \Sigma^i$  be a uniform  $\varepsilon$ -max-min  $\mathcal{D}_*^i$ -Markovian strategy, and let  $N_0 \in \mathbb{N}$  be the constant given in Definition 5.1. Then for every initial state  $s_0$ , every strategy profile  $\sigma^{-i} \in \Sigma^{-i}$  of the other players, and every  $k \geq 0$ , on the set  $\{\mathbf{E}_{s_0, \sigma^i, \sigma^{-i}}[\tau_{k+1}^{\mathcal{D}_*^i} - \tau_k^{\mathcal{D}_*^i} \mid \mathcal{H}(\tau_k^{\mathcal{D}_*^i})] \geq N_0\}$  we have*

$$\gamma(s_0, \sigma^i, \sigma^{-i}; \tau_k^{\mathcal{D}_*^i}, \tau_{k+1}^{\mathcal{D}_*^i}) \geq \underline{v}_1^i(s_{\tau_k^{\mathcal{D}_*^i}}) - \varepsilon.$$

## 6 The Modified Stochastic Game: The Max-Min Value

In this section we compare the limit as the discount factor goes to 1 of the max-min value in the modified game to the uniform max-min value in the original stochastic game.

The max-min value of player  $i$  in the modified game  $\widehat{\Gamma}_\lambda(s_0; \vec{\mathcal{D}}, \vec{c})$  is

$$\widehat{v}_\lambda^i(s_0; \mathcal{D}^i, c^i) := \max_{\sigma^i \in \Sigma^i} \min_{\sigma^{-i} \in \Sigma^{-i}} \widehat{\gamma}_\lambda^i(s_0, \sigma; \mathcal{D}^i, c^i).$$

Since the payoff function  $\sigma \mapsto \widehat{\gamma}_\lambda^i(s_0, \sigma; \mathcal{D}^i, c^i)$  is continuous (Lemma 4.7), the max-min value exists. By Eq. (5) we have  $\widehat{\gamma}_\lambda^i(s_0, \sigma; \mathcal{D}^i, c^i) \leq \gamma_\lambda^i(s_0, \sigma)$  for every strategy profile  $\sigma \in \Sigma$ , and consequently

$$\widehat{v}_\lambda^i(s_0; \mathcal{D}^i, c^i) \leq v_\lambda^i(s_0). \quad (19)$$

We do not know whether the function  $\lambda \mapsto \widehat{v}_\lambda^i(s_0; \mathcal{D}^i, c^i)$  is semi-algebraic, and in particular we do not know whether the limit  $\lim_{\lambda \rightarrow 1} \widehat{v}_\lambda^i(s_0; \mathcal{D}^i, c^i)$  always exists. If the limit exists, we denote it by  $\widehat{v}_1^i(s_0; \mathcal{D}^i, c^i)$  and term it the *limit max-min value* of player  $i$  in the modified game.

As the following example shows, the inequality in Eq. (19) can be strict, even if the cutoff vector  $c^i$  is high, and the difference between  $v_\lambda^i(s_0)$  and  $\widehat{v}_\lambda^i(s_0; \mathcal{D}^i, c^i)$  need not vanish as  $\lambda$  goes to 1.

**Example 6.1** *Consider the game that is depicted in Figure 2. In this game there is a single player, two states  $S = \{s^0, s^1\}$ , and the player has a single action in each state. Set  $\mathcal{D} = \{\{s^0\}, \{s^1\}\}$  and  $c(D) = 4$  for every  $D \in \mathcal{D}$ .*

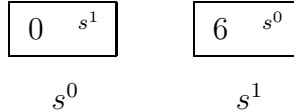


Figure 2: The game in Example 6.1.

*In this game the payoff alternates between 0 and 6, so that for each initial state the  $\lambda$ -discounted value of this Markov chain converges to 3 as the discount factor  $\lambda$*

goes to 1. In particular, the cutoff vector is higher than the  $\lambda$ -discounted value of the game for every discount factor  $\lambda$  sufficiently close to 1. In the modified game the payoff in state  $s^1$  is 4, so that the value of the modified game converges to 2 as  $\lambda$  goes to 1.

We will now provide conditions that ensure that the limit max-min value of a player in the modified game exists and coincides with his uniform max-min value in the original game. To this end we present a property, called *Property P*, that some partitions satisfy.

## 6.1 Property P

We turn to formally present Property P. Roughly, a partition  $\mathcal{D}^i$  satisfies Property P w.r.t. player  $i$  if for every  $\varepsilon > 0$  there is a uniform  $\varepsilon$ -max-min  $\mathcal{D}^i$ -Markovian strategy of player  $i$  such that the number of times in which the play switches between elements of  $\mathcal{D}^i$  is uniformly bounded, over the other players' strategy profile and over  $\varepsilon$ .

**Definition 6.2** *Let  $i \in I$  be a player. A partition  $\mathcal{D}$  of the set of states  $S$  satisfies Property P w.r.t. player  $i$  if there is a real number  $\zeta > 0$  and for every  $\varepsilon > 0$  there is a uniform  $\varepsilon$ -max-min  $\mathcal{D}$ -Markovian strategy  $\sigma^i \in \Sigma^i$  such that for every strategy profile  $\sigma^{-i} \in \Sigma^{-i}$  of the other players there is an event  $E$  for which  $\mathbf{P}_{s_0, \sigma^i, \sigma^{-i}}(E) > 1 - \varepsilon$  and*

$$\mathbf{E}_{s_0, \sigma^i, \sigma^{-i}}[Z(\cdot; \mathcal{D}) \cdot \mathbf{1}_E] \leq \zeta.$$

**Example 6.3 (Absorbing Games)** *When the game is an absorbing game, the state variable can change at most once, hence the element of the partition that contains the current state changes at most once. It follows that any partition  $\mathcal{D}$  satisfies Property P w.r.t. all players, with  $\zeta = 1$ .*

**Example 6.4** *Similarly to Example 6.3, suppose that there is a set  $D \subseteq S$  such that (a) all states that are not in  $D$  are absorbing, and (b)  $\mathcal{D} = \{D, \{s\}_{s \notin D}\}$ . Then the partition  $\mathcal{D}$  satisfies Property P w.r.t. all players, with  $\zeta = 1$ . In the application in Section 10 we will use this partition.*

**Example 6.5 (The partition  $\mathcal{D}_*^i$ )** *The partition  $\mathcal{D}_*^i$  of the set of states according to the uniform max-min value of player  $i$  satisfies Property P w.r.t. player  $i$ . Indeed, by Theorems 5.4 and 5.9, for every  $\varepsilon > 0$  player  $i$  has a  $\mathcal{D}_*^i$ -Markovian uniform  $\varepsilon$ -max-min strategy  $\sigma^i$ , and therefore by Theorem 5.7 this strategy satisfies Property P.*

## 6.2 A Result in Probability

We would like to prove that when the partition  $\mathcal{D}^i$  satisfies Property P, and the cutoffs  $(c^i(D))_{D \in \mathcal{D}^i}$  are sufficiently high, then the limit max-min value of a player in the modified game exists and coincides with his uniform max-min value in the original game. To this end we will show that a uniform  $\varepsilon$ -max-min  $\mathcal{D}^i$ -Markovian strategy  $\sigma^i$  of player  $i$  guarantees to him in the modified game at least  $\underline{v}_1^i(s_0) - 2\varepsilon$ . By Lemma 5.10, and because the strategy  $\sigma^i$  is uniform  $\varepsilon$ -max-min  $\mathcal{D}^i$ -Markovian, in every visit to an element of  $\mathcal{D}^i$  whose expected length is high, the expected average payoff is at least  $\underline{v}_1^i(s_0) - \varepsilon$ . Property P assures that the expected number of runs is bounded, hence the expected number of stages that belong to short visits to elements of  $\mathcal{D}^i$  is uniformly bounded. In particular, most stages belong to long visits to elements of  $\mathcal{D}^i$ . To show that the payoff in the modified game is high, we need to show that for every element  $D$  of  $\mathcal{D}^i$ , the discounted payoff during long visits to  $D$  is high. In this section we provide a result that will help us deduce that when the expected average payoff during long visits to  $D$  is high, so is the expected discounted payoff.

Let  $(\Omega, \mathcal{F}, P)$  be a probability space, let  $(\mathcal{F}_n)_{n \geq 0}$  be a filtration, and let  $(\tau_k)_{k \geq 0}$  be a sequence of stopping times adapted to the filtration  $(\mathcal{F}_n)_{n \geq 0}$  such that  $\tau_0 = 0$  and  $\tau_{k+1} > \tau_k$  on the event  $\{\tau_k < \infty\}$ , for every  $k \geq 0$ . Denote  $k_* := \sup_{k \geq 0} \{k \geq 0 : \tau_k < \infty\}$ .

For every nonnegative integer  $l$  denote by  $k(l)$  the unique nonnegative integer that satisfies  $\tau_{k(l)} \leq l < \tau_{k(l)+1}$ , and denote  $\tau_k^l := \tau_k \wedge l = \min\{\tau_k, l\}$ .

**Theorem 6.6** *In the notations above, If there is  $\zeta > 0$  such that  $\mathbf{E}[k_*] \leq \zeta$ , then for every  $\varepsilon > 0$  and every  $N \geq 0$  there are  $M = M(\varepsilon, N, \zeta) \geq 0$  and  $L_0 =$*

$L_0(\varepsilon, N, \zeta) \geq 0$  (both are independent of the probability measure  $P$ ) that satisfy

$$\sum_{l=0}^L P(\{\mathbf{E}[(\tau_{k(l)+1}^L - \tau_{k(l)}^L) \wedge M \mid \mathcal{F}_{\tau_{k(l)}}] \geq N\}) \geq (1 - \varepsilon)(L + 1), \quad \forall L \geq L_0.$$

For every  $k \geq 0$  call the sequence of stages between stages  $\tau_k$  and  $\tau_{k+1} - 1$  a  $\tau$ -run. The condition that  $\mathbf{E}[k_*]$  is finite assures that on average there are few  $\tau$ -runs up to stage  $L$ , hence, when we restrict attention to stages  $\{0, 1, 2, \dots, L\}$ , the expected length of  $\tau$ -runs is large. Theorem 6.6 states a stronger property: there is  $M$  such that the expected length of most  $\tau$ -runs, restricted to their first  $M$  stages, is high. Moreover, the constant  $M$  and the finite horizon  $L_0$  are independent of the probability measure  $P$ . In our application, the probability measure  $P$  will be determined by the strategies of the players. Since we study the max-min value, we need the quantities  $M$  and  $L_0$  to be independent of the strategies of the other player, hence  $M$  and  $L_0$  need to be independent of  $P$ .

The next example shows that the quantity  $M$  in Theorem 6.6 can be large when  $\zeta$  is large.

**Example 6.7** Let  $p \in (0, 1)$ , and consider a coin with parameter  $P(\text{head}) = p$  that is drawn over and over. For every  $k \geq 0$  let  $\tau_k$  be the following stopping time:

$$\tau_k := \begin{cases} k, & \text{if the first } k \text{ draws are head,} \\ \infty, & \text{otherwise.} \end{cases}$$

The first draw in which the outcome of the draw is tail is  $k_*$  (recall that time starts at 0), hence  $P(k_* = k) = (1 - p)p^k$ , and in particular  $\mathbf{E}[k_*] = \frac{1}{1-p}$ . Additionally, on the event  $\{\tau_k < \infty\}$  we have  $\mathbf{E}[(\tau_{k+1} - \tau_k) \wedge \frac{N-p}{1-p} \mid \tau_k < \infty] = N$ . It follows that the constant  $M$  in Theorem 6.6 may be at least  $\frac{N-p}{1-p}$ .

**Proof of Theorem 6.6.** We will substitute  $M = \lceil \frac{2\zeta N}{\varepsilon} \rceil$  and  $L_0 = \lceil \frac{2\zeta M}{\varepsilon} \rceil$ , where  $\lceil x \rceil$  is the least integer larger than or equal to the real number  $x$ . Denote

$$G_k := \{\tau_k < \infty\} \cap \{\tau_{k+1} - \tau_k < M\}.$$

Then

$$\zeta \geq \mathbf{E}[k_*] = \sum_{k=0}^{\infty} P(\tau_k < \infty) \geq \sum_{k=0}^{\infty} P(G_k). \quad (20)$$

Since  $\tau_{k+1}^L - \tau_k^L \leq \tau_{k+1} - \tau_k$ , and since on  $G_k$  we have  $\tau_{k+1} - \tau_k < M$ , it follows that for every  $L \geq 0$ ,

$$\mathbf{E} \left[ \sum_{k=0}^{\infty} (\tau_{k+1}^L - \tau_k^L) \mathbf{1}_{G_k} \right] < M \sum_{k=0}^{\infty} P(G_k) \leq \zeta M. \quad (21)$$

Denote

$$E_k := \{\tau_k < \infty\} \cap \{\mathbf{E}[(\tau_{k+1} - \tau_k) \wedge M \mid \mathcal{F}_{\tau_k}] < N\}.$$

By Markov's Inequality, on  $E_k$  we have

$$P(\tau_{k+1} - \tau_k < M \mid \mathcal{F}_{\tau_k}) = P((\tau_{k+1} - \tau_k) \wedge M < M \mid \mathcal{F}_{\tau_k}) > 1 - \frac{N}{M} \geq 1 - \frac{\varepsilon}{2\zeta}.$$

This implies that

$$P(E_k \cap G_k) \geq (1 - \frac{\varepsilon}{2\zeta})P(E_k),$$

which subsequently implies that

$$P(E_k \setminus G_k) \leq \frac{\varepsilon}{2\zeta} P(E_k). \quad (22)$$

As in Eq. (20) we have  $\sum_{k=0}^{\infty} P(E_k) \leq \zeta$ , and therefore by Eq. (22)

$$\sum_{k=0}^{\infty} P(E_k \setminus G_k) \leq \frac{\varepsilon}{2\zeta} \sum_{k=0}^{\infty} P(E_k) \leq \frac{\varepsilon}{2}.$$

Since  $\sum_{k=0}^{\infty} (\tau_{k+1}^L - \tau_k^L) = L + 1$ ,

$$\mathbf{E} \left[ \sum_{k=0}^{\infty} (\tau_{k+1}^L - \tau_k^L) \mathbf{1}_{E_k \setminus G_k} \right] \leq \frac{\varepsilon}{2} (L + 1). \quad (23)$$

From Eqs. (21) and (23) we obtain that

$$\begin{aligned}
L + 1 &= \mathbf{E} \left[ \sum_{k=0}^{\infty} (\tau_{k+1}^L - \tau_k^L) \right] \\
&= \mathbf{E} \left[ \sum_{k=0}^{\infty} (\tau_{k+1}^L - \tau_k^L) \mathbf{1}_{(E_k)^c} \right] + \mathbf{E} \left[ \sum_{k=0}^{\infty} (\tau_{k+1}^L - \tau_k^L) \mathbf{1}_{E_k \cap G_k} \right] + \mathbf{E} \left[ \sum_{k=0}^{\infty} (\tau_{k+1}^L - \tau_k^L) \mathbf{1}_{E_k \setminus G_k} \right] \\
&\leq \mathbf{E} \left[ \sum_{k=0}^{\infty} (\tau_{k+1}^L - \tau_k^L) \mathbf{1}_{(E_k)^c} \right] + \mathbf{E} \left[ \sum_{k=0}^{\infty} (\tau_{k+1}^L - \tau_k^L) \mathbf{1}_{G_k} \right] + \mathbf{E} \left[ \sum_{k=0}^{\infty} (\tau_{k+1}^L - \tau_k^L) \mathbf{1}_{E_k \setminus G_k} \right] \\
&\leq \mathbf{E} \left[ \sum_{k=0}^{\infty} (\tau_{k+1}^L - \tau_k^L) \mathbf{1}_{(E_k)^c} \right] + \zeta M + \frac{\varepsilon}{2} (L + 1) \\
&\leq \mathbf{E} \left[ \sum_{k=0}^{\infty} (\tau_{k+1}^L - \tau_k^L) \mathbf{1}_{(E_k)^c} \right] + \varepsilon (L + 1).
\end{aligned}$$

The result follows since

$$\mathbf{E} \left[ \sum_{k=0}^{\infty} (\tau_{k+1}^L - \tau_k^L) \mathbf{1}_{(E_k)^c} \right] = \sum_{l=0}^L P(\{\mathbf{E}[(\tau_{k(l)+1}^L - \tau_{k(l)}^L) \wedge M \mid \mathcal{F}_{\tau_{k(l)}}] \geq N\}).$$

■

### 6.3 Bounding the Max-Min Value in the Modified Game

Our goal in this section is to show that when the partition  $\mathcal{D}^i$  satisfies Property P w.r.t. player  $i$  and the cutoff  $c^i(D)$  is at least the uniform max-min value in all states in  $D$ , for every element  $D \in \mathcal{D}^i$ , then the limit max-min value of player  $i$  in the modified game exists and is equal to the uniform max-min value of the initial state.

**Theorem 6.8** *Let  $\mathcal{D}^i$  be a partition of the set of states  $S$  that satisfies Property P w.r.t. player  $i$  and let  $c^i \in \mathbb{R}^{\mathcal{D}^i}$  satisfy  $c^i(D) \geq \underline{v}_1^i(s)$  for every element  $D \in \mathcal{D}^i$  and every state  $s \in D$ . Then for every initial state  $s_0 \in S$  the limit  $\widehat{\underline{v}}_1^i(s_0; \mathcal{D}^i, c^i) := \lim_{\lambda \rightarrow 1} \widehat{\underline{v}}_\lambda^i(s_0; \mathcal{D}^i, c^i)$  exists we have  $\widehat{\underline{v}}_1^i(s_0; \mathcal{D}^i, c^i) = \underline{v}_1^i(s_0)$ .*



To prove Theorem 6.8 we will show that any uniform  $\varepsilon$ -max-min  $\mathcal{D}^i$ -Markovian strategy  $\sigma_{\mathcal{D},\varepsilon}^i$  guarantees approximately  $v_1^i(s_0)$  in the modified game  $\widehat{\Gamma}_\lambda(s_0; \mathcal{D}, c)$ , for every discount factor  $\lambda$  sufficiently close to 1.

For the proof of the theorem we will need two notations and an observation. For every discount factor  $\lambda \in [0, 1)$ , every strategy profile  $\sigma$ , and every two bounded stopping times  $0 \leq \tau \leq \tau'$  denote the expected unnormalized discounted time between the stopping times  $\tau$  and  $\tau'$  by

$$t_\lambda(s_0, \sigma; \tau, \tau') := \mathbf{E}_{s_0, \sigma} \left[ \sum_{n=\tau}^{\tau'-1} \lambda^{n-\tau} \mid \mathcal{H}(\tau) \right].$$

Denote the expected normalized discounted payoff between the stopping times  $\tau$  and  $\tau'$  by

$$\gamma_\lambda^i(s_0, \sigma; \tau, \tau') := \frac{\mathbf{E}_{s_0, \sigma} \left[ \sum_{n=\tau}^{\tau'-1} \lambda^{n-\tau} u^i(s_n, a_n) \mid \mathcal{H}(\tau) \right]}{t_\lambda(s_0, \sigma; \tau, \tau')}. \quad (24)$$

The quantities  $t_\lambda(s_0, \sigma; \tau, \tau')$  and  $\gamma_\lambda^i(s_0, \sigma; \tau, \tau')$  are random variables that depend on the information at stage  $\tau$ .

As is well known, the discounted payoff can be presented as a convex combination of the average payoffs. We will use the following relation, which relates the partial discounted sum to arithmetic sums, and holds for every  $\lambda \in [0, 1)$ , every  $L \in \mathbb{N}$ , every sequence of real numbers  $(x_n)_{n=0}^L$ , and every  $M \leq L$ :

$$\sum_{n=0}^L \lambda^n x_n = \sum_{n=0}^{M-1} (\lambda^n - \lambda^M) x_n + \sum_{l=M}^{\infty} \left( ((L \wedge l) + 1) (\lambda^l - \lambda^{l+1}) \frac{\sum_{n=0}^{L \wedge l} x_n}{(L \wedge l) + 1} \right). \quad (25)$$

Moreover, when  $M$  is fixed, the ratio  $\frac{\sum_{n=0}^{M-1} (\lambda^n - \lambda^M)}{\sum_{n=0}^L \lambda^n}$  of the sum of the coefficients in the first term on the right-hand side of Eq. (25) and the sum of the coefficients on the left-hand side goes to 0 as  $\lambda$  goes to 1. When replacing the nonnegative integer  $L$  by a stopping time  $\tau$ , Eq. (25) translates to

$$\mathbf{E} \left[ \sum_{n=0}^{\tau} \lambda^n x_n \right] = \mathbf{E} \left[ \sum_{n=0}^{(\tau \wedge M)-1} (\lambda^n - \lambda^{\tau \wedge M}) x_n \right] + \sum_{l=M}^{\infty} (\lambda^l - \lambda^{l+1}) \mathbf{E} [((\tau \wedge l) + 1)] \frac{\mathbf{E} [\sum_{n=0}^{\tau \wedge l} x_n]}{\mathbf{E} [(\tau \wedge l) + 1]}. \quad (26)$$

Moreover, the ratio  $\frac{\mathbf{E}\left[\sum_{n=0}^{(\tau \wedge M)-1} (\lambda^n - \lambda^{\tau \wedge M})\right]}{\mathbf{E}\left[\sum_{n=0}^{\tau} \lambda^n\right]}$  goes to 0 as  $\lambda$  goes to 1.

**Proof.** In view of Eq. (19) we need to show that  $\widehat{v}_1^i(s_0; \mathcal{D}^i, c^i) \geq \underline{v}_1^i(s_0)$ . Fix  $\delta > 0$ . Let  $\zeta$  be the constant of Definition 6.2, let  $\varepsilon > 0$  be the constant given by Theorem 5.9, let  $\sigma^i$  be the uniform  $\varepsilon$ -max-min  $\mathcal{D}^i$ -Markovian strategy for player  $i$  given by Property P, and let  $\sigma^{-i}$  be any strategy profile of the other players. We will show that the strategy  $\sigma^i$  guarantees that player  $i$ 's payoff in the modified game  $\widehat{\Gamma}_\lambda(s_0; \mathcal{D}^i, c^i)$  is at least  $\underline{v}_1^i(s_0) - 5\varepsilon - 2\delta$ , provided the discount factor is sufficiently close to 1.

Let  $N_0$  be the constant given by Theorem 5.4, and let  $M$  and  $L_0$  be the constant given by Theorem 6.6 for  $\zeta$  and  $N = N_0$ . Let  $\lambda$  be sufficiently close to 1 such that the ratio  $\frac{\mathbf{E}\left[\sum_{n=0}^{(\tau \wedge M)-1} (\lambda^n - \lambda^{\tau \wedge M})\right]}{\mathbf{E}\left[\sum_{n=0}^{\tau} \lambda^n\right]}$  is smaller than  $\varepsilon$ .

We argue that on the event  $\left\{ \mathbf{E}_{s_0, \sigma_{\mathcal{D}^i, \varepsilon}^i, \sigma^{-i}}[\tau_{k+1}^{\mathcal{D}} - \tau_k^{\mathcal{D}} \mid \mathcal{H}(\tau_k^{\mathcal{D}})] \geq N_0 \right\}$  we have

$$\gamma_\lambda^i(s_0, \sigma; \tau_k^{\mathcal{D}}, \tau_{k+1}^{\mathcal{D}} - 1) \geq \underline{v}_1^i(s_{\tau_k^{\mathcal{D}}}) - 3\varepsilon. \quad (27)$$

Indeed, on this event

$$\begin{aligned} & \gamma_\lambda^i(s_0, \sigma; \tau_k^{\mathcal{D}^i}, \tau_{k+1}^{\mathcal{D}^i} - 1) \\ &= \frac{\mathbf{E}_{s_0, \sigma} \left[ \sum_{n=\tau_k^{\mathcal{D}^i}}^{\tau_{k+1}^{\mathcal{D}^i}-1} \lambda^{n-\tau_k^{\mathcal{D}^i}} u^i(s_n, a_n) \mid \mathcal{H}(\tau_k^{\mathcal{D}^i}) \right]}{t_\lambda(s_0, \sigma; \tau_k^{\mathcal{D}^i}, \tau_{k+1}^{\mathcal{D}^i})} \end{aligned} \quad (28)$$

$$= \frac{\mathbf{E} \left[ \sum_{n=0}^{(\tau_k^{\mathcal{D}^i} \wedge M)-1} (\lambda^n - \lambda^{\tau \wedge M}) u^i(s_n, a_n) \mid \mathcal{H}(\tau_k^{\mathcal{D}^i}) \right]}{t_\lambda(s_0, \sigma; \tau_k^{\mathcal{D}^i}, \tau_{k+1}^{\mathcal{D}^i})} \quad (29)$$

$$\begin{aligned} &+ \frac{\sum_{l=M}^{\infty} (\lambda^l - \lambda^{l+1}) \mathbf{E}[(\tau_k^{\mathcal{D}^i} \wedge l) + 1] \gamma^i(s_0, \sigma^i, \sigma^{-i}; \tau_k^{\mathcal{D}^i}, \tau_{k+1}^{\mathcal{D}^i})}{t_\lambda(s_0, \sigma; \tau_k^{\mathcal{D}^i}, \tau_{k+1}^{\mathcal{D}^i})} \\ &\geq \sum_{l=M}^{\infty} (\lambda^l - \lambda^{l+1}) \mathbf{E}[(\tau_k^{\mathcal{D}^i} \wedge l) + 1] \gamma^i(s_0, \sigma^i, \sigma^{-i}; \tau_k^{\mathcal{D}^i}, \tau_{k+1}^{\mathcal{D}^i}) - \varepsilon \end{aligned} \quad (30)$$

$$\geq (\underline{v}_1^i(s_0) - \varepsilon) \sum_{l=M}^{\infty} (\lambda^l - \lambda^{l+1}) \mathbf{E}[(\tau_k^{\mathcal{D}^i} \wedge l) + 1] - \varepsilon \quad (31)$$

$$\geq (\underline{v}_1^i(s_0) - \varepsilon)(1 - \varepsilon) - \varepsilon \quad (32)$$

$$\geq \underline{v}_1^i(s_0) - 3\varepsilon, \quad (33)$$

where Eq. (28) holds by definition (see Eq. (24)), Eq. (29) holds by Eq. (26), Eq. (30) holds by the choice of  $\lambda$ , Eq. (31) holds by the choice of  $M$ , and Eq. (32) holds by the choice of  $\lambda$ .

Finally, the payoff in the modified game satisfies

$$\begin{aligned} & \widehat{\gamma}_\lambda^i(s_0, \sigma^i, \sigma^{-i}; \mathcal{D}^i, c^i) \\ &= \mathbf{E}_{s_0, \sigma^i, \sigma^{-i}} \left[ \sum_{D \in \mathcal{D}^i} \min \{ U_\lambda^i(s_0, \sigma^i, \sigma^{-i}; D), t_\lambda(s_0, \sigma^i, \sigma^{-i}; D) \cdot c^i(D) \} \right] \end{aligned} \quad (34)$$

$$\geq \sum_{D \in \mathcal{D}^i} \sum_{k=0}^{\infty} \mathbf{E}_{s_0, \sigma^i, \sigma^{-i}} \left[ \mathbf{1}_{\{\tau_k^{\mathcal{D}^i} < \infty\} \cap \{D(s_{\tau_k^{\mathcal{D}^i}}) = D\}} \cdot \lambda^{\tau_k^{\mathcal{D}^i} - 1} \cdot \right. \quad (35)$$

$$\begin{aligned} & \left. t_\lambda(s_0, \sigma; \tau_k^{\mathcal{D}^i}, \tau_{k+1}^{\mathcal{D}^i}) \min \{ \gamma_\lambda^i(s_0, \sigma^i, \sigma^{-i}; \tau_k^{\mathcal{D}^i}, \tau_{k+1}^{\mathcal{D}^i}), c^i(D) \} \right] \\ & \geq \sum_{D \in \mathcal{D}^i} \sum_{k=0}^{\infty} \mathbf{E}_{s_0, \sigma^i, \sigma^{-i}} \left[ \mathbf{1}_{\{\tau_k^{\mathcal{D}^i} < \infty\} \cap \{D(s_{\tau_k^{\mathcal{D}^i}}) = D\}} t_\lambda(s_0, \sigma^i, \sigma^{-i}; \tau_k^{\mathcal{D}^i}, \tau_{k+1}^{\mathcal{D}^i}) \cdot (\underline{v}_1^i(s_{\tau_k^{\mathcal{D}^i}}) - 3\varepsilon) \right] \\ & \quad - 2\varepsilon \end{aligned} \quad (36)$$

$$\geq \underline{v}_1^i(s_0) - 5\varepsilon - 2\delta. \quad (37)$$

where Eq. (34) holds by the definition of the modified payoff, Eq. (35) follows holds since

$$\min\{a_1, b_1\} + \min\{a_2, b_2\} \leq \min\{a_1 + a_2, b_1 + b_2\}, \quad \forall a_1, a_2, b_1, b_2 \in \mathbb{R}, \quad (38)$$

Eq. (36) holds by Theorem 6.6, Eq. (27), and since  $c^i(D) \geq \underline{v}_1^i(s)$  for every element  $D \in \mathcal{D}^i$  and every state  $s \in D$ , and Eq. (37) holds since  $\sigma^i$  is  $\delta$ -max-min preserving. ■

## 7 The Modified Game: the Min-Max Value

In Section 6 we studied the max-min value in the modified game. In the present section we derive the analogous results with regards to the min-max value. The proofs of these results are similar to the proofs regarding the max-min value, hence omitted. The reason we started with the study of the max-min value is that while player  $i$  has one strategy that guarantees that his payoff is at least his max-min value minus  $\varepsilon$ , whatever the other players play, this is not the case for the min-max value: for every strategy profile of the other players, player  $i$  may have a different

strategy that guarantees that his payoff is at least the uniform min-max value minus  $\varepsilon$ . This feature of the concept of the min-max value has the implication that definitions and proofs are more cumbersome, yet they pose no new technical difficulties.

The  $\lambda$ -discounted min-max value of player  $i$  at the initial state  $s_0$  in the original stochastic game is given by

$$\bar{v}_\lambda^i(s_0) := \min_{\sigma^{-i} \in \Sigma^{-i}} \max_{\sigma^i \in \Sigma^i} \gamma_\lambda^i(s_0; \sigma^i, \sigma^{-i}). \quad (39)$$

The interpretation of the min-max value is that the other players can ensure that player  $i$ 's payoff will not be above his min-max value, and they cannot lower his payoff further. Denote

$$\bar{v}_1^i(s_0) := \lim_{\lambda \rightarrow 1} \bar{v}_\lambda^i(s_0),$$

which is termed the *uniform min-max value* of player  $i$  at the initial state  $s_0$ .

Neyman (2003) proved that for every strategy profile of players  $I \setminus \{i\}$  player  $i$  has a response that guarantees that his payoff is at least the uniform min-max value minus  $\varepsilon$ , provided the discount factor is sufficiently close to 1 or the game is sufficiently long. We will need this strategy to satisfy an additional condition, which resembles the concept of being Markovian w.r.t. a partition. This condition is complicated to formulate, since the strategy profile of players  $I \setminus \{i\}$  need not be Markovian w.r.t. the partition. We therefore spell out directly the properties that the response of player  $i$  should satisfy.

Denote by  $\mathcal{D}_{**}^i$  the partition of the set of states  $S$  according to the uniform min-max value of player  $i$ : two states  $s, s' \in S$  are in the same element of the partition if and only if  $\bar{v}_1^i(s) = \bar{v}_1^i(s')$ . The next result follows from Neyman (2003) together with the analog of Theorem 5.9.

**Theorem 7.1** *For every  $\varepsilon > 0$ , every player  $i \in I$ , and every initial state  $s_0 \in S$  there is  $N_0 \in \mathbb{N}$  such that for every strategy profile  $\sigma^{-i} \in \Sigma^{-i}$  there is a strategy  $\sigma^i$  such that for every  $N \geq N_0$ ,*

$$\gamma_N^i(s_0, \sigma^i, \sigma^{-i}) \geq \bar{v}_1^i(s_0) - \varepsilon. \quad (40)$$

Moreover, for every  $k \geq 0$  and every bounded stopping time  $\tau \geq \tau_k^{\mathcal{D}^{i**}}$ ,

$$\bar{v}_1^i(s_{\tau_k^{\mathcal{D}^{i**}}}) \leq \mathbf{E}_{s_{\tau_k^{\mathcal{D}^{i**}}}, \sigma_h^i, \sigma_h^{-i}}^{\mathcal{D}^{i**}} [\bar{v}_1^i(s_\tau)] + \varepsilon,$$

and on the event  $\left\{ \mathbf{E}_{s_0, \sigma^i, \sigma^{-i}}[\tau - \tau_k^{\mathcal{D}^{i**}} \mid \mathcal{H}(\tau_k^{\mathcal{D}^{i**}})] \geq N_0 \right\}$  we have

$$\gamma^i(s_0, \sigma^i, \sigma^{-i}; \tau_k^{\mathcal{D}^{i**}}, \tau) \geq \bar{v}_1^i(s_{\tau_k^{\mathcal{D}^{i**}}}) - \varepsilon. \quad (41)$$

Note that Eq. (41) requires some sort of uniform subgame perfectness: the strategy profile  $\sigma^i$  is a good response in every subgame that starts when the play switches an element of the partition  $\mathcal{D}_{**}^i$ . We will call a strategy  $\sigma^i$  that satisfies Eqs. (40)–(41) for the partition  $\mathcal{D}_{**}^i$  a *uniform  $\mathcal{D}_{**}^i$ -subgame-perfect  $\varepsilon$ -min-max strategy against  $\sigma^{-i}$* .

The analog of Property P for the min-max value is the following.

**Definition 7.2** *Let  $i \in I$  be a player. A partition  $\mathcal{D}$  of the set of states  $S$  satisfies Property P' w.r.t. player  $i$  if there is a real number  $\zeta > 0$  and for every  $\varepsilon > 0$  and every strategy profile  $\sigma^{-i} \in \Sigma^{-i}$  there are a uniform  $\mathcal{D}$ -subgame-perfect  $\varepsilon$ -min-max strategy  $\sigma^i \in \Sigma^i$  against  $\sigma^{-i}$  and an event  $E$  such that  $\mathbf{P}_{s_0, \sigma^i, \sigma^{-i}}(E) \geq 1 - \varepsilon$  and*

$$\mathbf{E}_{s_0, \sigma^i, \sigma^{-i}}[Z \cdot \mathbf{1}_E] \leq \zeta.$$

The partitions in Examples 6.3 and 6.4 satisfy Property P' w.r.t. all players. Analogously to Example 6.5, for every player  $i \in I$  the partition  $\mathcal{D}_{**}^i$  satisfies Property P' w.r.t. player  $i$ .

Denote by

$$\widehat{v}_\lambda^i(s_0; \mathcal{D}^i, c^i) := \min_{\sigma^{-i} \in \Sigma^{-i}} \max_{\sigma^i \in \Sigma^i} \gamma_\lambda^i(s_0; \sigma^i, \sigma^{-i})$$

the min-max value of player  $i$  in the modified game  $\widehat{\Gamma}_\lambda^i(s_0; \vec{\mathcal{D}}, \vec{c})$ . The following result, which is analogous to Theorem 6.8, states that if the partition  $\mathcal{D}^i$  satisfies Property P' and if the vector  $c^i \in \mathbb{R}^{\mathcal{D}^i}$  is at least the min-max value, then the limit min-max value of player  $i$  in the modified game is his uniform min-max value at the initial state in the original stochastic game.

**Theorem 7.3** *Let  $\mathcal{D}^i$  be a partition of the set of states  $S$  that satisfies Property  $P'$  w.r.t. player  $i$  and let  $c^i \in \mathbb{R}^{\mathcal{D}^i}$  satisfy  $c^i(D) \geq \bar{v}_1^i(s)$  for every element  $D \in \mathcal{D}^i$  and every state  $s \in D$ . Then for every initial state  $s_0 \in S$  the limit  $\hat{v}_1^i(s_0; \mathcal{D}^i, c^i) := \lim_{\lambda \rightarrow 1} \hat{v}_\lambda^i(s_0; \mathcal{D}^i, c^i)$  exists and satisfies  $\hat{v}_1^i(s_0; \mathcal{D}^i, c^i) = \bar{v}_1^i(s_0)$ .*

When particularized to absorbing games, Theorem 7.3 coincides with Step 2 in the proof of Theorem 4.5 in Solan (1999).

Since the payoff to a player in an equilibrium is always at least his min-max value, Theorem 7.3 yields that the limit of stationary equilibrium payoffs of each player  $i$  in the modified game is at least his uniform min-max value in the original stochastic game at the initial state.

**Corollary 7.4** *Fix an initial state  $s_0 \in S$ , a vector of partitions  $\vec{\mathcal{D}} = (\mathcal{D}^i)_{i \in I}$  of the set of states such that for every player  $i \in I$  the partition  $\mathcal{D}^i$  satisfies Property  $P'$  w.r.t. player  $i$ , and a vector  $\vec{c} = (c^i(D))_{i \in I, D \in \mathcal{D}^i} \in \mathbb{R}^{I \times \mathcal{D}}$  such that  $c^i(D) \geq \bar{v}_1^i(s)$  for every player  $i \in I$ , every element  $D \in \mathcal{D}^i$ , and every state  $s \in D$ . Let  $\lambda \mapsto x_\lambda$  be a function that assigns to every discount factor  $\lambda \in [0, 1)$  a stationary equilibrium in the modified game  $\hat{\Gamma}_\lambda(s_0; \vec{\mathcal{D}}, \vec{c})$ . Then  $\liminf_{\lambda \rightarrow 1} \hat{\gamma}_\lambda^i(s_0, x_\lambda; \mathcal{D}^i, c^i) \geq \bar{v}_1^i(s_0)$ .*

## 8 The Min-Max Value in Stationary Strategies

In discounted stochastic games the min-max (resp. max-min) value of a player when all players are restricted to stationary strategies coincides with his min-max (resp. max-min) value without this restriction. In this section (resp. Section 9) we study the min-max (resp. max-min) value in the modified game when the players are restricted to stationary strategies, and check whether the above mentioned phenomenon holds in the modified game. The motivation for this study is that if one can prove that there exists a stationary strategy profile  $x_\lambda$  that guarantees that the payoff of each player in the modified game is at least his min-max value (resp. max-min value), whatever be the initial state, and if the limit min-max value (resp. max-min value) in stationary strategies in the modified game when

the players are restricted to stationary strategies coincides with his uniform min-max value (resp. uniform max-min value) in the original stochastic game, then we may be able to derive results on the original stochastic game using the sequence  $(x_\lambda)_{\lambda \in [0,1]}$ .

The min-max value of player  $i$  in the modified game  $\hat{\Gamma}_\lambda(s_0; \vec{\mathcal{D}}, \vec{c})$  when the players are restricted to stationary strategies is

$$\hat{\bar{v}}_{\lambda, \text{stat}}^i(s_0; \mathcal{D}^i, c^i) := \min_{x^{-i} \in \Sigma_{\text{stat}}^{-i}} \max_{x^i \in \Sigma_{\text{stat}}^i} \hat{\gamma}_\lambda^i(s_0, x^i, x^{-i}; \mathcal{D}^i, c^i). \quad (42)$$

Since the payoff function  $\hat{\gamma}_\lambda^i(s_0, \cdot; \mathcal{D}^i, c^i) : \Sigma \rightarrow \mathbb{R}$  is continuous, the min-max value in stationary strategies is well-defined. In Eq. (42), the maximum and minimum are over the set of stationary strategies of the players, hence both are smaller than the sets considered for the calculation of the min-max value  $\bar{v}_\lambda^i(s_0; \mathcal{D}^i, c^i)$ . Consequently, there is no clear comparison between the min-max value and the min-max value in stationary strategies of the modified game. Nevertheless we have the following result.

**Theorem 8.1** *For every discount factor  $\lambda \in [0, 1)$ , every initial state  $s_0 \in S$ , every player  $i \in I$ , every partition  $\mathcal{D}^i$  of the set of states, and every vector  $c^i \in \mathbb{R}^{\mathcal{D}^i}$ ,*

$$\hat{\bar{v}}_\lambda^i(s_0; \mathcal{D}^i, c^i) \leq \hat{\bar{v}}_{\lambda, \text{stat}}^i(s_0; \mathcal{D}^i, c^i) \leq \bar{v}_\lambda^i(s). \quad (43)$$

**Proof.** The left-hand side inequality in Eq. (43) holds due to the following chain of equalities and inequalities:

$$\hat{\bar{v}}_{\lambda, \text{stat}}^i(s_0; \mathcal{D}^i, c^i) = \min_{x^{-i} \in \Sigma_{\text{stat}}^{-i}} \max_{x^i \in \Sigma_{\text{stat}}^i} \hat{\gamma}_\lambda^i(s_0, x^i, x^{-i}; \mathcal{D}^i, c^i) \quad (44)$$

$$= \min_{x^{-i} \in \Sigma_{\text{stat}}^{-i}} \max_{\sigma^i \in \Sigma^i} \hat{\gamma}_\lambda^i(s_0, \sigma^i, x^{-i}; \mathcal{D}^i, c^i) \quad (45)$$

$$\geq \min_{\sigma^{-i} \in \Sigma^{-i}} \max_{\sigma^i \in \Sigma^i} \hat{\gamma}_\lambda^i(s_0, \sigma^i, \sigma^{-i}; \mathcal{D}^i, c^i) \quad (46)$$

$$= \hat{\bar{v}}_\lambda^i(s_0; \mathcal{D}^i, c^i), \quad (47)$$

where Eq. (45) follows from Theorem 4.2.



We turn to prove the right-hand side inequality in Eq. (43). Because the discounted min-max value in the original stochastic game is attained in stationary strategies and by Eq. (5) we have

$$\begin{aligned}\bar{v}_\lambda^i(s_0) &= \min_{x^{-i} \in \Sigma_{\text{stat}}^{-i}} \max_{x^i \in \Sigma_{\text{stat}}^i} \gamma_\lambda^i(s_0, x^i, x^{-i}) \\ &\geq \min_{x^{-i} \in \Sigma_{\text{stat}}^{-i}} \max_{x^i \in \Sigma_{\text{stat}}^i} \hat{\gamma}_\lambda^i(s_0, x^i, x^{-i}; \mathcal{D}^i, c^i) \\ &= \hat{\bar{v}}_{\lambda, \text{stat}}^i(s_0; \mathcal{D}^i, c^i),\end{aligned}$$

as claimed. ■

By taking the limits as  $\lambda$  goes to 1 in Eq. (43) and by using Theorem 7.3 we obtain the following result.

**Corollary 8.2** *For every initial state  $s_0 \in S$ , every player  $i \in I$ , every partition  $\mathcal{D}^i$  of the set of states that satisfies Property P' w.r.t. player  $i$ , and every vector  $c^i \in \mathbb{R}^{\mathcal{D}^i}$  that satisfies that  $c^i(D) \geq \bar{v}_1^i(s)$  for every element  $D \in \mathcal{D}^i$  and every state  $s \in D$ , we have*

$$\lim_{\lambda \rightarrow 1} \hat{\bar{v}}_{\lambda, \text{stat}}^i(s_0; \mathcal{D}^i, c^i) = \lim_{\lambda \rightarrow 1} \hat{\bar{v}}_\lambda^i(s_0; \mathcal{D}^i, c^i).$$

## 9 The Max-Min Value in Stationary Strategies

In this section we define the concept of max-min value in stationary strategies in the modified game, and show that its behavior is different than the behavior of the min-max value in stationary strategies.

The max-min value of player  $i$  in the modified game  $\hat{\Gamma}_\lambda(s_0; \vec{\mathcal{D}}, \vec{c})$  when the players are restricted to stationary strategies is

$$\hat{\underline{v}}_{\lambda, \text{stat}}^i(s_0; \mathcal{D}^i, c^i) := \max_{x^i \in \Sigma_{\text{stat}}^i} \min_{x^{-i} \in \Sigma_{\text{stat}}^{-i}} \hat{\gamma}_\lambda^i(s_0, x^i, x^{-i}; \mathcal{D}^i, c^i).$$

We first show that the max-min value is not lowered when the other players are allowed to play any strategy profile.

**Theorem 9.1** *For every discount factor  $\lambda \in [0, 1)$ , every initial state  $s_0 \in S$ , every player  $i \in I$ , every partition  $\mathcal{D}^i$  of the set of states, and every vector  $c^i \in \mathbb{R}^{\mathcal{D}^i}$ ,*

$$\hat{v}_{\lambda, \text{stat}}^i(s_0; \mathcal{D}^i, c^i) = \max_{x^i \in \Sigma_{\text{stat}}^i} \min_{\sigma^{-i} \in \Sigma^{-i}} \hat{\gamma}_{\lambda}^i(s_0, x^i, \sigma^{-i}; \mathcal{D}^i, c^i).$$

Theorem 9.1 follows from the following result, which states that when player  $i$  plays a stationary strategy, the lowest payoff of player  $i$  in the modified game is obtained when players  $I \setminus \{i\}$  play a pure stationary strategy profile.

**Lemma 9.2** *For every discount factor  $\lambda \in [0, 1)$ , every initial state  $s_0 \in S$ , every player  $i \in I$ , every partition  $\mathcal{D}^i$  of the set of states, every vector  $c^i \in \mathbb{R}^{\mathcal{D}^i}$ , every stationary strategy  $x^i \in \Sigma_{\text{stat}}^i$ , and every strategy profile  $\sigma^{-i} \in \Sigma^{-i}$  of the other players there is a pure stationary strategy profile  $x^{-i} \in \Sigma_{\text{stat}}^{-i}$  such that*

$$\hat{\gamma}_{\lambda}^i(s_0, x^i, \sigma^{-i}; \mathcal{D}^i, c^i) \geq \hat{\gamma}_{\lambda}^i(s_0, x^i, x^{-i}; \mathcal{D}^i, c^i).$$

**Proof.** When player  $i$  plays a stationary strategy and the other players are conceived as a single decision maker who can correlate its  $|I| - 1$  actions, the decision problem of players  $I \setminus \{i\}$  becomes a Markov decision problem. By Altman (1999, Theorem 3.2) there are  $L \in \mathbb{N}$ , a probability distribution  $\beta \in \Delta(\{1, 2, \dots, L\})$ , and a collection of pure stationary strategy profiles  $(x_l^{-i})_{l=1}^L$  of all players other than player  $i$  such that

$$t_{\lambda}(s_0, x^i, \sigma^{-i}) = \sum_{l=1}^L \beta_l t_{\lambda}(s_0, x^i, x_l^{-i}).$$

By Eq. (2), for every element  $D \in \mathcal{D}^i$ ,

$$\begin{aligned} U_{\lambda}^i(s_0, x^i, \sigma^{-i}; D) &= \sum_{s \in D, a \in A(s)} t_{\lambda}(s_0, x^i, \sigma^{-i}; s, a) u^i(s, a) \\ &= \sum_{s \in D, a \in A(s)} \sum_{l=1}^L \beta_l t_{\lambda}(s_0, x^i, x_l^{-i}) u^i(s, a) \\ &= \sum_{l=1}^L \beta_l U_{\lambda}^i(s_0, x^i, x_l^{-i}; D). \end{aligned}$$

By Eq. (38) we have

$$\begin{aligned}
\widehat{\gamma}_\lambda^i(s_0, x^i, \sigma^{-i}; \mathcal{D}^i, c^i) &= \sum_{D \in \mathcal{D}^i} \min \{U_\lambda^i(s_0, x^i, \sigma^{-i}; D), t_\lambda(s_0, x^i, \sigma^{-i}; D) \cdot c^i(D)\} \\
&= \sum_{D \in \mathcal{D}^i} \min \left\{ \sum_{l=1}^L \beta_l U_\lambda^i(s_0, x^i, x_l^{-i}; D), \sum_{l=1}^L \beta_l t_\lambda(s_0, x^i, x_l^{-i}; D) \cdot c^i(D) \right\} \\
&\geq \sum_{D \in \mathcal{D}^i} \sum_{l=1}^L \beta_l \min \{U_\lambda^i(s_0, x^i, x_l^{-i}; D), t_\lambda(s_0, x^i, x_l^{-i}; D) \cdot c^i(D)\} \\
&= \sum_{l=1}^L \beta_l \sum_{D \in \mathcal{D}^i} \min \{U_\lambda^i(s_0, x^i, x_l^{-i}; D), t_\lambda(s_0, x^i, x_l^{-i}; D) \cdot c^i(D)\} \\
&= \sum_{l=1}^L \beta_l \widehat{\gamma}_\lambda^i(s_0, x^i, x_l^{-i}; \mathcal{D}^i, c^i) \\
&\geq \min_{l=1, \dots, L} \widehat{\gamma}_\lambda^i(s_0, x^i, x_l^{-i}; \mathcal{D}^i, c^i),
\end{aligned}$$

and the desired result follows. ■

We can now prove that the max-min value in stationary strategies in the modified game is not larger than the max-min value in the modified game.

**Theorem 9.3** *For every discount factor  $\lambda \in [0, 1)$ , every initial state  $s_0 \in S$ , every player  $i \in I$ , every partition  $\mathcal{D}^i$  of the set of states, and every vector  $c^i \in \mathbb{R}^{\mathcal{D}^i}$ ,*

$$\widehat{\underline{v}}_{\lambda, \text{stat}}^i(s_0; \mathcal{D}^i, c^i) \leq \widehat{\underline{v}}_\lambda^i(s_0; \mathcal{D}^i, c^i). \quad (48)$$

**Proof.** The claim follows from the following chain of equalities and inequality, that holds due to Lemma 9.2.

$$\begin{aligned}
\widehat{\underline{v}}_\lambda^i(s_0; \mathcal{D}^i, c^i) &= \max_{\sigma^i \in \Sigma^i} \min_{\sigma^{-i} \in \Sigma^{-i}} \gamma_\lambda^i(s_0; \sigma^i, \sigma^{-i}) \\
&\geq \max_{x^i \in \Sigma_{\text{stat}}^i} \min_{\sigma^{-i} \in \Sigma^{-i}} \gamma_\lambda^i(s_0; x^i, \sigma^{-i}) \\
&= \max_{x^i \in \Sigma_{\text{stat}}^i} \min_{x^{-i} \in \Sigma_{\text{stat}}^{-i}} \gamma_\lambda^i(s_0; x^i, \sigma^{-i}) \\
&= \widehat{\underline{v}}_{\lambda, \text{stat}}^i(s_0; \mathcal{D}^i, c^i).
\end{aligned}$$

■

The following example shows that the inequality in Theorem 9.3 can be strict and not vanish as the discount factor goes to 1.

**Example 9.4 (The Big Match)** *The two-player zero-sum absorbing game that appears in Figure 3 is a slight variation of the Big Match, which was introduced by Gillette (1956) and analyzed by Blackwell and Ferguson (1968).*

	$L$	$R$								
$T$	<table><tr><td><math>0, 1</math></td><td><math>s^2</math></td></tr></table>	$0, 1$	$s^2$	<table><tr><td><math>1, 0</math></td><td><math>s^1</math></td></tr></table>	$1, 0$	$s^1$				
$0, 1$	$s^2$									
$1, 0$	$s^1$									
$B$	<table><tr><td><math>0, 1</math></td><td><math>s^0</math></td></tr></table>	$0, 1$	$s^0$	<table><tr><td><math>1, 0</math></td><td><math>s^0</math></td></tr></table>	$1, 0$	$s^0$	$B$	<table><tr><td><math>0, 1</math></td><td><math>s^1</math></td></tr></table>	$0, 1$	$s^1$
$0, 1$	$s^0$									
$1, 0$	$s^0$									
$0, 1$	$s^1$									
		$s^0$		$s^1$						
				$s^2$						

Figure 3: The game in Example 9.4.

Let  $\mathcal{D}^1$  be the partition of  $S$  in which each state forms a distinct element:

$$\mathcal{D}^1 := \{\{s^0\}, \{s^1\}, \{s^2\}\}.$$

Let  $c^1$  be the vector composed of the uniform value of Player 1 in the various states:

$$c^1(\{s^0\}) := \frac{1}{2}, \quad c^1(\{s^1\}) := 0, \quad c^1(\{s^2\}) := 1.$$

By Blackwell and Ferguson (1968) and Theorem 6.8,

$$\widehat{v}_1^1(s_0; \mathcal{D}^1, c^1) = \underline{v}_1^1(s_0) = \frac{1}{2}.$$

We will show that the inequality in Theorem 9.3 can be strict by showing that for every discount factor  $\lambda \in [0, 1)$  we have

$$\widehat{v}_{\lambda, \text{stat}}^1(s^0; \mathcal{D}^1, c^1) = \frac{1}{3}.$$

Any stationary strategy  $[p(T), (1-p)(B)]$  of Player 1 determines the total  $\lambda$ -discounted time  $\alpha$  that the play spends in  $s^0$  before it is absorbed. This  $\lambda$ -discounted time is denoted by  $\alpha = \alpha_\lambda(p)$  and is given by

$$\begin{aligned} \alpha_\lambda(p) &= (1-\lambda) + \lambda(1-\lambda)(1-p) + \lambda^2(1-\lambda)(1-p)^2 + \cdots \\ &= \frac{1-\lambda}{1-\lambda(1-p)}. \end{aligned} \tag{49}$$

The function  $p \mapsto \alpha_\lambda(p)$  given in Eq. (9.4) is a monotone decreasing function from  $[0, 1]$  to  $[1 - \lambda, 1]$  and onto. It follows that

$$\begin{aligned} \hat{v}_{\lambda, stat}^1(s^0; \mathcal{D}^1, c^1) &= \max_{x^1 \in \Sigma_{stat}^1} \min_{x^2 \in \Sigma_{stat}^2} \hat{\gamma}_\lambda^1(s_0, x^1, x^2; \mathcal{D}^1, c^1) \\ &= \max_{\alpha \in [1-\lambda, 1]} \min_{y \in [0, 1]} (\alpha \cdot \min\{1 - y, \frac{1}{2}\} + (1 - \alpha)y) \\ &= \frac{1}{3}, \end{aligned} \quad (50)$$

and the maximum in Eq. (50) is attained at  $\alpha = \frac{2}{3}$ .

## 10 Application: Uniform Equilibrium in Strongly Controllable Games

In this section we show that the modified game can be used to prove the existence of a uniform  $\varepsilon$ -equilibrium in a certain class of multiplayer stochastic games. We start by describing the class of games that we study. Call a set of states  $D$  *closed* if the play cannot leave  $D$ , whatever the players play. Call the set  $D$  *strongly controllable* if the play can leave  $D$  only when the play is in a specific state  $s_D \in D$ , and a specific player  $i_D$  plays a specific action  $\tilde{a}^{i_D}$ ; in all other states in  $D$  the play cannot leave  $D$ , and when the play is in state  $s_D$  and player  $i_D$  plays an action that is not  $\tilde{a}^{i_D}$ , the play cannot leave  $D$ .

Let  $\mathcal{D}$  be the partition of the set of states according to the vector of uniform min-max values of the players: two states are in the same element of the partition if all players have the same uniform min-max values in the two states. Consider the following refinement  $\mathcal{D}^*$  of  $\mathcal{D}$ : two states  $s$  and  $s'$  are in the same element of  $\mathcal{D}^*$  if (a) they lie in the same element of  $\mathcal{D}$ , (b) there is a strategy profile that ensures that the play reaches  $s'$  without leaving  $D$  when the initial state is  $s$ , and (c) there is a strategy profile that ensure that the play reaches  $s$  without leaving  $D$  when the initial state is  $s'$ . We call a stochastic game *strongly controllable* if each element of the partition  $\mathcal{D}^*$  is either closed or strongly controllable.

We will prove that every strongly controllable multiplayer stochastic game admits a uniform equilibrium payoff. Admittedly, this class of games is restricted,

yet the proof that is described below and shows that games in this class admit a uniform  $\varepsilon$ -equilibrium, which involves the modified game, is the first to achieve this task. An alternative way to prove the existence of a uniform  $\varepsilon$ -equilibrium in this class of games uses the technique of Solan and Vieille (2002, Section 4).

Strongly controllable games are by no means significant in their own right. The goal of this section is to show a proof of concept: the modified game can be used to prove the existence of a uniform equilibrium payoff in some nontrivial class of multiplayer stochastic games. We hope that in the future more convincing applications to this tool will be found.

**Definition 10.1** *Let  $\varepsilon \geq 0$ . A strategy profile  $\sigma$  is a uniform  $\varepsilon$ -equilibrium if for every initial state  $s_0 \in S$ , every player  $i \in I$ , every integer  $N \in \mathbb{N}$  sufficiently large, and every discount factor  $\lambda \in [0, 1)$  sufficiently close to 1,*

$$\gamma_\lambda^i(s_0, \sigma'^i, \sigma^{-i}) \leq \gamma_\lambda^i(s_0, \sigma) + \varepsilon, \quad \forall \sigma'^i \in \Sigma^i,$$

and

$$\gamma_N^i(s_0, \sigma'^i, \sigma^{-i}) \leq \gamma_N^i(s_0, \sigma) + \varepsilon, \quad \forall \sigma'^i \in \Sigma^i.$$

The existence of a uniform  $\varepsilon$ -equilibrium has been verified for various classes of stochastic games (see, e.g., Mertens and Neyman (1981), Solan (1999), Vieille (2000a,b), Solan and Vieille (2001), Flesch, Thuijsman, and Vrieze (2007), Simon (2007, 2012), and Flesch, Schoenmakers, and Vrieze (2008, 2009)). In this section we concentrate on the class of strongly controllable stochastic games, which we define now formally.

**Definition 10.2** *A set of states  $D \subseteq S$  is closed if  $q(D \mid s, a) = 1$  for every state  $s \in D$  and every action profile  $a \in A(s)$ . A set of states  $D \subseteq S$  is strongly controllable if the following condition holds: there is a player  $i_D \in I$ , a state  $s_D \in D$ , and an action  $\tilde{a}^{i_D} \in A^{i_D}(s_D)$  such that for every state  $s \in D$  and every action profile  $a \in A(s)$ , if  $q(D \mid s, a) < 1$  then  $s = s_D$  and  $a^{i_D} = \tilde{a}^{i_D}$ .*

Let  $\mathcal{D}$  be the partition of the set of states according to the uniform min-max value vector: two states  $s, s' \in S$  are in the same element of  $\mathcal{D}$  if and only if  $\bar{v}_1^i(s) = \bar{v}_1^i(s')$  for every player  $i \in I$ .

For every set  $C \subseteq S$  denote by  $\nu_C$  the first arrival time to  $C$ :

$$\nu_C := \min\{n \geq 0: s_n \in C\}.$$

By convention, the minimum of an empty set is  $+\infty$ .

**Definition 10.3** *Let  $D \subseteq S$  be a set of states and let  $s, s' \in D$ . We say that state  $s$  leads in  $D$  to state  $s'$  if there is a strategy profile  $\sigma$  such that under  $\sigma$  the play reaches  $s'$  before leaving  $D$  when the initial state is  $s$ :*

$$\mathbf{P}_{s,\sigma}(\nu_{\{s'\}} < \nu_{S \setminus D}) = 1. \quad (51)$$

*The states  $s$  and  $s'$  are  $D$ -siblings if each one leads in  $D$  to the other.*

If there is a strategy profile  $\sigma$  that satisfies Eq. (51), then there is a pure stationary strategy profile that satisfies this equation, and, moreover, this strategy profile is independent of  $s$ , as soon as  $s \in D$ . By definition, every state leads to itself in any set  $D$  that contains it. As a result, the  $D$ -siblings relation is reflexive, commutative, and transitive. The following result states that if the states  $s$  and  $s'$  are  $D$ -siblings, and if  $s''$  is a state that is visited with positive probability under the strategy that leads in  $D$  from  $s$  to  $s'$ , then the states  $s'$  and  $s''$  are  $D$ -siblings as well.

**Lemma 10.4** *Let  $D \subseteq S$  be a set of states, let  $s, s' \in D$ , and let  $\sigma$  be a strategy profile that satisfies Eq. (51). Let  $s'' \in D$  be a state that satisfies  $\mathbf{P}_{s,\sigma}(\nu_{\{s''\}} < \nu_{\{s'\}}) > 0$ . If  $s$  and  $s'$  are  $D$ -siblings, then  $s'$  and  $s''$  are  $D$ -siblings.*

**Proof.** The part of  $\sigma$  from its first visit to  $s''$  until the play reaches  $s'$  is a strategy profile that leads from  $s''$  to  $s'$  without leaving  $D$ .

We now prove that there is a strategy profile that ensures that the play reaches  $s''$  without leaving  $D$  when the initial state is  $s'$ . Let  $\sigma'$  be a strategy profile that ensures that the play reaches state  $s$  without leaving  $D$  when the initial state is  $s'$ . Let  $\sigma''$  be the strategy profile in which the players alternately follow  $\sigma'$  until the play reaches  $s$  and then they follow  $\sigma$  until the play reaches  $s'$ . By definition,

under  $\sigma''$  the play never leaves  $D$ , that is,  $\mathbf{P}_{s',\sigma''}(\nu_{S \setminus D} = \infty) = 1$ . We claim that under  $\sigma''$  the play reaches  $s''$  with probability 1, that is,  $\mathbf{P}_{s',\sigma''}(\nu_{\{s''\}} < \infty) = 1$ . Indeed, after each visit to  $s$ , with probability  $\mathbf{P}_{s,\sigma}(\nu_{\{s''\}} < \nu_{\{s'\}})$  the play reaches state  $s''$  before reaching state  $s'$ , hence the play eventually reaches state  $s''$  a.s. ■

Two states  $s, s'$  are  $\mathcal{D}$ -siblings if they are  $D$ -siblings for some element  $D \in \mathcal{D}$ . Denote by  $\mathcal{D}^*$  the partition of the set of states according to the  $\mathcal{D}$ -siblings equivalence relation. This partition is a refinement of the partition  $\mathcal{D}$ . Lemma 10.4 implies that if the states  $s$  and  $s'$  lie in the same element of the partition  $\mathcal{D}^*$ , then state  $s$  leads to state  $s'$  without leaving the element of  $\mathcal{D}^*$  that contains both states  $s$  and  $s'$ .

**Definition 10.5** *A stochastic game is strongly controllable if all elements  $D \in \mathcal{D}^*$  are either closed or strongly controllable.*

Closed sets of states resemble repeated games, since the play cannot leave such a set, can move between the states, and the min-max value is independent of the state. Consequently, whenever the initial state is in a closed set of states, a uniform  $\varepsilon$ -equilibrium exists. In fact, in this case a folk theorem obtains, and every feasible and individually rational payoff vector is a uniform equilibrium payoff, namely a limit as  $\varepsilon$  goes to 0 of payoff vectors that correspond to uniform  $\varepsilon$ -equilibria. The main result of this section is the following.

**Theorem 10.6** *Every strongly controllable stochastic game  $\Gamma$  admits a uniform  $\varepsilon$ -equilibrium, for every  $\varepsilon > 0$ .*

The rest of the section is devoted to the proof of Theorem 10.6, which uses the approach of Vrieze and Thuijsman (1989) and Solan (1999), and, except of the use of the modified game, is by now standard. We will therefore provide the main lines of the proof without getting into all the details.

Fix  $\varepsilon > 0$  throughout, and for the time being also fix an element  $D \in \mathcal{D}^*$ .

**Step 1:** Definition of a restricted stochastic game and its modified game.



Consider the stochastic game  $\Gamma^D$  that is derived from the stochastic game  $\Gamma$  by restricting the set of states to  $D$ ; that is,

- The set of players is  $I$ .
- The set of states is  $S$ ; all states  $s \notin D$  are absorbing with absorbing payoff  $\bar{v}_1(s)$ .
- The set of actions of each player  $i$  in each state  $s \in D$  is  $A^i(s)$ .
- Transitions and stage payoff in states in  $D$  are as in the original stochastic game.

If  $D$  is a closed set of states, let  $s_D$  be an arbitrary state in  $D$ ; if  $D$  is a strongly controllable set, the state  $s_D$  was defined in Definition 10.2. Define a partition  $\mathcal{D}_D$  over the set of states  $S$  by

$$\mathcal{D}_D := \{D, \{s\}_{s \notin D}\},$$

and for every player  $i \in I$  define a vector of cut-offs  $(c^i(E))_{E \in \mathcal{D}_D}$  by

$$c_D^i(D) := \bar{v}_1^i(s_D), \quad c_D^i(\{s\}) := \bar{v}_1^i(s) \quad \forall s \notin D.$$

Since the set  $D$  contains all nonabsorbing states in the game  $\Gamma^D$ , the partition  $\mathcal{D}_D$  satisfies Property P' w.r.t. every player with  $\zeta = 1$  (see the paragraph after Definition 7.2).

Consider the modified game  $\hat{\Gamma}_\lambda^D(s_D; \mathcal{D}_D, \vec{c}_D)$  that is based on the restricted stochastic game  $\Gamma^D$ , where all players share the same partition  $\mathcal{D}_D$ , and the vector of cutoffs is  $\vec{c} := (c_D^i)_{i \in I}$ . For every  $\lambda \in [0, 1)$  let  $x_\lambda$  be a stationary equilibrium in the modified game  $\hat{\Gamma}_\lambda^D(s_D; \mathcal{D}_D, \vec{c}_D)$  such that the function  $\lambda \mapsto x_\lambda$  is semi-algebraic (see Theorem 4.11). In particular the limit  $x_1(s) := \lim_{\lambda \rightarrow 1} x_\lambda(s)$  exists for every state  $s \in S$ , and by Corollary 7.4 we have

$$\lim_{\lambda \rightarrow 1} \hat{\gamma}_\lambda^i(s_0, x_\lambda; \mathcal{D}_D, c_D^i) \geq \bar{v}_1^i(s_D). \quad (52)$$

The payoff in the modified game  $\widehat{\Gamma}_\lambda^D(s_D; \mathcal{D}_D, \vec{c}_D)$  under the strategy profile  $x_\lambda$  is

$$\begin{aligned} \widehat{\gamma}_\lambda^{D,i}(s_D, x_\lambda; \mathcal{D}_D, \vec{v}_1^i) &= \sum_{E \in \mathcal{D}} \min \{U_\lambda^i(s_D, x_\lambda; E), t_\lambda(s_D, x_\lambda; E) \cdot \vec{c}_D^i(E)\} \\ &= \min \{U_\lambda^i(s_D, x_\lambda; D), t_\lambda(s_D, x_\lambda; D) \cdot \vec{v}_1^i(s_D)\} + \sum_{s \notin D} t_\lambda(s_D, x_\lambda; \{s\}) \cdot \vec{v}_1^i(s). \end{aligned} \quad (53)$$

By taking the limit as  $\lambda$  goes to 1 in Eq. (53) and using Eq. (52) we deduce that exactly one of the following alternatives hold:

- (A.1)  $\lim_{\lambda \rightarrow 1} t_\lambda(s_D, x_\lambda; D) = 1$  and  $\lim_{\lambda \rightarrow 1} U_\lambda^i(s_D, x_\lambda; D) \geq \vec{v}_1^i(s_D)$  for every player  $i \in I$ .
- (A.2)  $\lim_{\lambda \rightarrow 1} t_\lambda(s_D, x_\lambda; D) < 1$  and  $\sum_{s \notin D} t_\lambda(s_D, x_\lambda; \{s\}) \cdot \vec{v}_1^i(s) \geq \vec{v}_1^i(s_D)$  for every player  $i \in I$ .

The significance of the modified game is in implying this dichotomy. We will show that if for  $D \in \mathcal{D}^*$  Condition (A.1) holds, then when the initial state is in  $D$  there is a uniform  $\varepsilon$ -equilibrium under which the play never leaves  $D$ , while if Condition (A.2) holds, then when the initial state is in  $D$  there is a uniform  $\varepsilon$ -equilibrium under which the play leaves  $D$  with probability 1.

**Step 2:** Condition (A.1) holds.

The condition  $\lim_{\lambda \rightarrow 1} t_\lambda(s_D, \sigma; D) = 1$  arises in two cases:

- The set  $D$  is closed, and in particular the play cannot leave it.
- The set  $D$  is strongly controllable, and as  $\lambda$  goes to 1 the  $\lambda$ -discounted time at which the play leaves  $D$  under  $x_\lambda$  goes to 0:  $\lim_{\lambda \rightarrow 1} \mathbf{E}_{s_0, x_\lambda}[(1 - \lambda)\lambda^{\nu_{s \setminus D}}] = 0$ .

Both cases are treated in the same way.

Since all states in the set  $D$  are  $D$ -siblings (Lemma 10.4), for every state  $s' \in D$  there is a stationary strategy profile that ensures that the play reaches state  $s'$  without leaving  $D$  when the initial state is in  $D$ . If  $t_\lambda(s_D, \sigma; \{s'\}) > 0$  for every  $\lambda \in [0, 1]$ , that is, the  $\lambda$ -discounted time spent in state  $s'$  is positive 0, then there is

a stationary strategy profile that is a perturbation of  $x_1$  and ensures that the play reaches  $s'$  without leaving  $D$  when the initial state is in  $D$ . Formally, for every  $\delta > 0$  there is a stationary strategy profile  $y_{\delta,s'}$  that satisfies

- The strategy profile  $y_{\delta,s'}$  is a perturbation of  $x_1$ , that is,  $\|y_{\delta,s'}^i(s) - x_1^i(s)\|_\infty < \delta$  for every player  $i \in I$  and every state  $s \in D$ .
- Under the strategy profile  $y_{\delta,s'}$  the play reaches state  $s'$  with probability 1 without leaving  $D$ , provided the initial state is in  $D$ :

$$\mathbf{P}_{s,y_{\delta,s'}}(\nu_{\{s'\}} < \nu_{S \setminus D}) = 1, \quad \forall s \in D.$$

Since  $\lim_{\lambda \rightarrow 1} t_\lambda(s_D, x_\lambda; D) = 1$ , Eq. (4) implies that

$$\lim_{\lambda \rightarrow 1} U_\lambda^i(s_D, x_\lambda; D) = \lim_{\lambda \rightarrow 1} \gamma_\lambda^i(s_D, x_\lambda),$$

hence Condition (A.1) implies that

$$\lim_{\lambda \rightarrow 1} \gamma_\lambda^i(s_D, x_\lambda) \geq \bar{v}_1^i(s_D), \quad \forall i \in I.$$

The quantity  $\lim_{\lambda \rightarrow 1} \gamma_\lambda(s_D, x_\lambda)$  is a convex combination of the payoff under the stationary strategy profile  $x_1$  in its irreducible sets. Recall that a nonempty set  $C \subseteq S$  is *irreducible under  $x_1$*  if

- (I.1) The set  $C$  is closed under  $x_1$ , that is,  $q(C \mid s, x_1) = 1$  for every state  $s \in C$ .
- (I.2) The set  $C$  is a minimal set (w.r.t. set inclusion) that satisfies property (I.1).

Denote by  $\mathcal{I}(D; x_1)$  the collection of all irreducible sets under  $x_1$  that are contained in  $D$ . It is well known that for every irreducible set  $C$  under  $x_1$  the limit  $\lim_{\lambda \rightarrow 1} \gamma_\lambda(s_0, x_1)$  is independent of the initial state  $s_0$ , as long as  $s_0$  is in  $C$ . We denote this limit by  $\gamma_1(C, x_1)$ . For every irreducible set  $C$  under  $x_1$  denote the  $\lambda$ -discounted time that the play spends in the set  $C$  under strategy profile  $x_\lambda$  when the initial state is  $s_D$  by

$$\beta(C) := \lim_{\lambda \rightarrow 1} t_\lambda(s_D, x_\lambda; C).$$

Then

$$\lim_{\lambda \rightarrow 1} \gamma_{\lambda}^i(s_D, x_{\lambda}) = \sum_{C \in \mathcal{I}(D; x_1)} \beta(C) \cdot \gamma_1(C, x_1).$$

Denote  $\text{supp } (\beta) = \{C_1, C_2, \dots, C_L\}$  and for each  $l \in \{1, 2, \dots, L\}$  choose a state  $s_l \in C_l$ .

Consider the following strategy profile  $\sigma_K$  that depends on a positive real number  $K$ :

1. Set  $l = 1$ .
2. The players play the stationary strategy profile  $y_{\varepsilon, s_l}$  until the play reaches a state in  $C_l$ .
3. The players play the stationary strategy profile  $x_1$  for  $\lceil \beta(C_l) \cdot K \rceil$  stages.
4. Increase  $l$  by 1 modulo  $L$  and go to Step 2.

The reader can verify that as  $K$  increases the payoff under  $\sigma_K$  converges to  $\sum_{C \in \mathcal{I}(D; x_1)} \beta(C) \cdot \gamma_1(C, x_1)$ , that is,

$$\lim_{K \rightarrow \infty} \lim_{\lambda \rightarrow 1} \gamma_{\lambda}^i(s_D, \sigma_K) = \sum_{C \in \mathcal{I}(D; x_1)} \beta(C) \cdot \gamma_1(C, x_1),$$

and

$$\lim_{K \rightarrow \infty} \lim_{N \rightarrow \infty} \gamma_N^i(s_D, \sigma_K) = \sum_{C \in \mathcal{I}(D; x_1)} \beta(C) \cdot \gamma_1(C, x_1).$$

It is standard to verify that by adding statistical tests to  $\sigma_K$ , which check whether the players indeed follow it, the strategy profile  $\sigma_K$  can be turned into a uniform  $2\varepsilon$ -equilibrium, provided the initial state is in  $D$ .

Choose  $K_0 \in \mathbb{N}$  sufficiently large such that

$$\left| \lim_{\lambda \rightarrow 1} \gamma_{\lambda}^i(s_D, \sigma_{K_0}) - \sum_{C \in \mathcal{I}(D; x_1)} \beta(C) \cdot \gamma_1(C, x_1) \right| \leq \varepsilon,$$

and

$$\left| \lim_{N \rightarrow \infty} \gamma_N^i(s_D, \sigma_{K_0}) - \sum_{C \in \mathcal{I}(D; x_1)} \beta(C) \cdot \gamma_1(C, x_1) \right| \leq \varepsilon.$$

**Step 3:** Condition (A.2) holds.

The condition  $\lim_{\lambda \rightarrow 1} t_\lambda(s_D, x_\lambda; D) < 1$  implies that the play can leave  $D$ , hence in particular the set  $D$  is not closed, and therefore it is strongly controllable. By Condition (A.2),

$$\sum_{s \notin D} \lim_{\lambda \rightarrow 1} t_\lambda(s_D, x_\lambda; \{s\}) \cdot \bar{v}_1^i(s) \geq \bar{v}_1^i(s_D), \quad \forall i \in I,$$

and therefore

$$\lim_{\lambda \rightarrow 1} \left( \frac{\sum_{s' \notin D} q(s' \mid s_D, \tilde{a}^{i_D}, x_\lambda^{-i_D}) \cdot \bar{v}_1^i(s')}{q(S \setminus D \mid s_D, \tilde{a}^{i_D}, x_\lambda^{-i_D})} \right) \geq \bar{v}_1^i(s_D), \quad \forall i \in I.$$

A strategy profile that ensures that the play leaves the set  $D$  and the expected continuation uniform min-max value is high suggests itself. Fix  $\lambda_0 \in [0, 1)$  that satisfies

$$\frac{\sum_{s' \notin D} q(s' \mid s_D, \tilde{a}^{i_D}, x_{\lambda_0}^{-i_D}) \cdot \bar{v}_1^i(s')}{q(S \setminus D \mid s_D, \tilde{a}^{i_D}, x_{\lambda_0}^{-i_D})} \geq \bar{v}_1^i(s_D) - \delta, \quad \forall i \in I,$$

where  $\delta > 0$  will be determined below.

- The players play a pure stationary strategy that guarantees that the play reaches the state  $s_D$ .
- At state  $s_D$  players  $I \setminus \{i_D\}$  play  $x_{\lambda_0}^{-i_D}$ , while player  $i_D$  plays  $(1 - \eta)\hat{a}^{i_D} + \eta\tilde{a}^{i_D}$ , where  $\hat{a}^{i_D}$  is some action in  $A^{i_D}(s_D) \setminus \{\tilde{a}^{i_D}\}$ , and  $\eta > 0$  is sufficiently small.
- Deviations in states in  $D \setminus \{s_D\}$  are observed immediately and punished at the deviator's uniform min-max level.
- Deviations in  $s_D$  of players in  $I \setminus \{i_D\}$  are observed statistically, provided  $\eta$  is sufficiently small.

- If the play has not left  $D$  after  $\frac{1}{\eta^2}$  visits to state  $s_D$  and no deviation has been detected, all players start punishing player  $i_D$  at his uniform min-max level. The constant  $\eta$  is chosen sufficiently small so that if no player deviates, the probability that the play does not leave  $D$  after  $\frac{1}{\eta^2}$  visits to state  $s_D$  is smaller than  $\delta$ .

**Step 4:** Combining the strategies.

Denote by  $\sigma^*$  the strategy profile that plays as follows:

- Whenever the play enters an element  $D \in \mathcal{D}^*$  that satisfies  $\lim_{\lambda \rightarrow 1} t_\lambda(s_D, x_\lambda; D) = 1$ , the strategy profile  $\sigma^*$  follows the strategy profile  $\sigma_{K_0}$  defined in Step 2.
- Whenever the play enters an element  $D \in \mathcal{D}^*$  that satisfies  $\lim_{\lambda \rightarrow 1} t_\lambda(s_D, x_\lambda; D) < 1$ , the strategy profile  $\sigma^*$  follows the strategy profile defined in Step 3 until the play leaves the set  $D$ .

The strategy profiles defined in Step 2 ensures that, when the initial state is in an element  $D \in \mathcal{D}^*$  that satisfies Condition (A.1), the discounted payoff and the average payoff of each player  $i$  is at least  $\bar{v}_1^i(s_D) - 2\delta$ , provided the discount factor is sufficiently close to 1 and the horizon is sufficiently long. The strategy profile defined in Step 3 ensures that, when the initial state is in an element  $D \in \mathcal{D}^*$  that satisfies Condition (A.2), the play leaves  $D$  and the expected continuation uniform min-max value of each player  $i$  is at least  $\bar{v}_1^i(s_D) - \delta$ . Note that for every player  $i \in I$ , the strategy  $\sigma^{*i}$  is  $\mathcal{D}_D$ -Markovian.

Recall that  $k(\mathcal{D}^*; n)$  is the number of times in which the play switched elements of the partition  $\mathcal{D}^*$  up to stage  $n$ . For each player  $i \in I$  define a stochastic process  $(W_n)_{n \geq 0}$  with values in  $\mathbb{R}^I$  by  $W_n^i := \bar{v}_1^i(s_n) + \delta \cdot k(\mathcal{D}^*; n)$  for each  $i \in I$  and every  $n \geq 0$ . By the construction of the strategy profile  $\sigma^*$ , the process  $(W_n^i)_{n \geq 0}$  is a bounded submartingale under the strategy profile  $\sigma^*$ , for every player  $i \in I$ .

We argue that the expected number of times in which the play switches between elements of  $\mathcal{D}^*$  is bounded by a constant that is independent of  $\delta$ . Indeed, the proof of Theorem 5.7 implies that the number of times in which the play switches

between elements of  $\mathcal{D}$  is bounded by  $C(\rho)$ . Moreover, if  $s, s'$  are two states in the same elements of  $\mathcal{D}$  but in different elements of  $\mathcal{D}^*$ , then one of them does not lead to the other. It follows that the expected number of times in which the process changes an element of the partition  $\mathcal{D}^*$  before leaving the current element of the partition  $\mathcal{D}$  is uniformly bounded by a constant  $C'$  that depends only on the transition function  $q$ . The claim follows.

As in the proof of Theorem 5.7 we deduce that there is an event  $E$  that satisfies the following conditions:

- $\mathbf{P}_{s_0, \sigma^*}(E) \geq 1 - \varepsilon$  for every initial state  $s_0$ .
- On the event  $E$  the number of times the process  $(W_k)_{k \geq 0}$  changes its values is uniformly bounded, say by  $C''$ , which depends on  $\rho$  and on the transition function  $q$ .

We choose  $\delta$  sufficiently small such that  $C''\delta \leq \varepsilon$ .

We deduce that  $\mathbf{P}_{s_0, \sigma^*}$ -a.s. the play reaches a closed element in  $\mathcal{D}^*$ , and therefore the expected long-run average payoff of each player  $i$  under  $\sigma^*$  is at least  $\bar{v}_1^i(s_0) - 3\varepsilon$ . Moreover, for every history  $h_n \in H$  the expected long-run average payoff of each player  $i$  under  $\sigma^*$ , conditioned that history  $h_n$  occurred, is at least  $\bar{v}_1^i(s_n) - 3\varepsilon$ . This in turn implies that the strategy profile  $\sigma^*$  is uniform  $4\varepsilon$ -equilibrium.

## 11 Discussion and Open Problems

In this paper we defined for every stochastic game an auxiliary game, which we termed the modified game, studied some of its properties, and showed that equilibria in this auxiliary game can be used to study uniform equilibrium in the original stochastic game. This approach is similar to the vanishing discount factor approach of Vrieze and Thuijsman (1989), who studied uniform equilibrium in two-player absorbing games by analyzing an auxiliary game, which, in their case, was the discounted game. Solan (1999) took this approach one step further by studying a more involved version of the payoff function in absorbing games, by altering

the payoff function in the nonabsorbing state. In the present paper we extended Solan's (1999) approach to general stochastic games. Vieille (2000b), Solan (2000), and Solan and Vieille (2002) studied another type of auxiliary game when the underlying game is a recursive game by restricting the players to completely mixed strategies.

Unlike Vrieze and Thuijsman (1989), Solan (1999), Vieille (2000b), Solan (2000), and Solan and Vieille (2002), the auxiliary game presented in this paper is valid for every stochastic game, and not only for absorbing games or for recursive games. It is probable that our modified game is not the only auxiliary game that can be useful to studying general stochastic games, and that in the future other types of auxiliary games will be proposed and studied.

Example 4.10 shows that there need not be a stationary strategy profile  $x$  that satisfies Eq. (6) for every player  $i \in I$  and every initial state  $s_0 \in S$ . We do not know whether for every  $\varepsilon > 0$  there exists a stationary strategy profile  $x$  that satisfies

$$\hat{\gamma}_\lambda^i(s_0, x; \mathcal{D}^i, c^i) \geq \bar{v}_1^i(s_0) - \varepsilon, \quad \forall i \in I, s_0 \in S.$$

The existence of such a stationary strategy profile may have significant implications on the study of acceptability in stochastic games (Solan, 2016) and on the study of uniform correlated  $\varepsilon$ -equilibrium in stochastic games (Solan and Vieille, 2002; Solan and Vohra, 2002).

As mentioned before, we do not know whether the functions  $\lambda \mapsto \hat{v}_\lambda^i(s_0; \mathcal{D}^i, c^i)$  and  $\lambda \mapsto \bar{v}_\lambda^i(s_0; \mathcal{D}^i, c^i)$  are semi-algebraic for every initial state  $s_0 \in S$ , every player  $i \in I$ , every partition  $\mathcal{D}^i$  of the set of states, and every vector  $c^i \in \mathbb{R}^{\mathcal{D}^i}$ . It is interesting to know whether this is indeed the case, and if not, whether the limits of these functions as the discount factor goes to 1 nevertheless exist.

## References

- [1] Altman E. (1999) *Constrained Markov Decision Processes*, Chapman and Hall/CRC.



- [2] Bewley T. and Kohlberg E. (1976) The Asymptotic Theory of Stochastic Games. *Mathematics of Operations Research*, **1**, 197–208.
- [3] Billingsley P. (1995) *Probability and Measure*, John Wiley & Sons.
- [4] Blackwell D. and Ferguson T. S. (1968) The Big Match. *The Annals of Mathematical Statistics*, **39**, 159–163.
- [5] Buhovsky L., Solan E., and Solan O.N. (2016) Monovex Sets, Preprint.
- [6] Eilenberg S. and Montgomery D. (1946) Fixed Point Theorems for Multi-Valued Transformations. *American Journal of Mathematics*, **68**, 214–222.
- [7] Fink A.M. (1964) Equilibrium in a Stochastic  $n$ -Person Game, *Journal of Science of the Hiroshima University*, Series A-I Math, **28**, 89–93.
- [8] Flesch J., Schoenmakers G., and Vrieze K. (2008) Stochastic Games on a Product State Space, *Mathematics of Operations Research*, **33**, 403–420.
- [9] Flesch J., Schoenmakers G., and Vrieze K. (2009) Stochastic Games on a Product State Space: the Periodic Case, *International Journal of Game Theory*, **38**, 263–289.
- [10] Flesch J., Thuijsman F. and O.J. Vrieze (1997) Stochastic Games with Additive Transitions, *European Journal of Operational Research*, **179**, 483–497.
- [11] Flesch J., Thuijsman F., and Vrieze O.J. (1997) Cyclic Markov Equilibria in Stochastic Games. *International Journal of Game Theory*, **26**, 303–314.
- [12] Flesch J., Thuijsman F., and Vrieze O.J. (2007) Stochastic Games with Additive Transitions. *European Journal of Operations Research*, **179**, 483–497.
- [13] Gillette D. (1957) Stochastic Games with Zero Stop Probabilities. *Contributions to the Theory of Games*, **3**, Princeton University Press

- [14] Hörner J., Sugaya T., Takahashi S., and Vieille N. (2011) Recursive Methods in Discounted Stochastic Games: an Algorithm for  $\delta \rightarrow 1$  and a Folk Theorem. *Econometrica*, **79**, 1277–1318.
- [15] Mertens J.F. and Neyman A. (1981) Stochastic Games, *International Journal of Game Theory*, **10**, 53–66.
- [16] Mertens J.F. and T. Parthasarathy (1987) Equilibria for Discounted Stochastic Games, CORE Discussion Paper No. 8750. Appeared in *Stochastic Games and Applications*, A. Neyman and S. Sorin (Eds.), Kluwer Academic Publishers, 131–172.
- [17] Neyman A. (2003) Real Algebraic Tools in Stochastic games. In *Stochastic Games and Applications*, Neyman A. and Sorin S. (eds.), Kluwer Academic Publishers, 57–75.
- [18] Simon R.S. (2007) The Structure of Non-Zero-Sum Stochastic Games, *Advances in Applied Mathematics*, **38**, 1–26.
- [19] Simon R.S. (2012) A Topological Approach to Quitting Games, *Mathematics of Operations Research*, **37**, 180–195.
- [20] Solan E. (1999) Three-Player Absorbing Games. *Mathematics of Operations Research*, **24**, 669–698.
- [21] Solan E. (2000) Absorbing Team Games. *Games and Economic Behavior*, **31**, 245–261.
- [22] Solan E. (2016) Acceptable Strategy Profiles in Stochastic Games. *Games and Economic Behavior*, Forthcoming.
- [23] Solan E. and Vieille N. (2001) Quitting Games. *Mathematics of Operations Research*, **26**, 265–285.
- [24] Solan E. and Vieille N. (2002) Correlated Equilibrium in Stochastic Games. *Games and Economic Behavior*, **38**, 362–399.

- [25] Solan E. and Vohra R. (2002) Correlated Equilibrium Payoffs and Public Signalling in Absorbing Games, *International Journal of Game Theory*, **31**, 91–121.
- [26] Takahashi M. (1964) Equilibrium Points of Stochastic Non-Cooperative  $n$ -Person Games, *Journal of Science of the Hiroshima University, Series A-I Math*, **28**, 95–99.
- [27] Vieille N. (2000a) Two-Player Stochastic Games I: A Reduction. *Israel Journal of Mathematics*, **119**, 55–91.
- [28] Vieille N. (2000b) Two-Player Stochastic Games II: The Case of Recursive Games. *Israel Journal of Mathematic*, **119**, 93–126.
- [29] Vrieze O.J. and Thuijsman F. (1989) On Equilibria in Repeated Games with Absorbing States. *International Journal of Game Theory*, **18**, 293–310.